VOLUME IV, No. 1 November, 1961

EDUCATIONAL RESEARCH

A Review for
Teachers and all concerned with
Progress in Education

SECRETARY'S OFFICE COPY
DO NOT REMOVE

Published for
THE NATIONAL FOUNDATION FOR EDUCATIONAL
RESEARCH IN ENGLAND AND WALES

by
NEWNES EDUCATIONAL PUBLISHING CO. LTD.
Tower House, Southampton Street, London, W.C.2

EDUCATIONAL RESEARCH

Editor:
W. D. WALL
assisted by
D. A. PIDGEON K. M. MILLER
D. WILSON BELL

Educational Research is published for the National Foundation by Newnes Educational Publishing Co., Ltd., once a term, in November, February and June. The price per issue (post free) throughout the world is 5/6 and the annual subscription (post free) is 16/6.

Subscriptions, orders and advertisements should be sent to Newnes Educational Publishing Co., Ltd., Tower House, Southampton Street, London, W.C.2.

Communications on general editorial matters including MSS. for publication, books and other printed matter for review or abstracting should be sent to the Director, N.F.E.R., 79, Wimpole Street, London, W.1.

Note to contributors: Contributions which make a readable synthesis of valid research on the practical problems posed by the English educational system will be particularly welcome. The length of such contributions should be from 4,000 to 7,000 words, though shorter notes will be considered. Abstracts of theses or similar material should be of sufficient length and detail to acquaint the reader with the main outcomes of the work, the sampling on which it has been based and the experimental statistical and other techniques used. Such synopses should be submitted in agreement with the head of the university department concerned.

All papers submitted for publication should conform to the conventional usage of the review and especially to the form in which bibliographical references are given. They should be typed (double space) on one side of quarto paper.

Intending contributors would be well advised in the first instance to send to the editor a brief synopsis of their paper with the elements of a bibliography.

In certain cases it may be possible for the Foundation's Information Service to aid in the gathering of relevant research material.

Only material which has not been published already elsewhere can normally be accepted for publication. The copyright in any material published is retained by the Editor. Reproduction in whole or in part is forbidden except (i) under the terms of the Fair Copying Declaration of the Royal Society or (ii) when specific permission has been given by the Editor.

Contributors receive fifteen off-prints of their article and five free copies of the number in which it appears. Further off-prints may be ordered at cost price when the proofs are returned to the Editor.

The conclusions drawn and the opinions expressed in papers are those of the authors. They should not be taken to represent the view of the N.F.E.R.

Order Form for Binder

TO

NEWNES EDUCATIONAL PUBLISHING CO., LTD.,
BOOK SALES DEPT.,
TOWER HOUSE, SOUTHAMPTON STREET,
LONDON, W.C.2

Dear Sirs,

Please supply one binder to take 9 issues (Vols. I, II and III)* of EDUCATIONAL RESEARCH. I enclose cheque/money order/postal order for 10/6d (plus postage 1/6d.)

NAME .....

ADDRESS .....

.....

.....

BLOCK CAPITALS PLEASE

.....

SIGNATURE

* Later supplies of the binder have no vol. nos.

ORDER FORM

To
NEWNES EDUCATIONAL PUBLISHING CO., LTD.,
BOOK SALES DEPT.,
TOWER HOUSE, SOUTHAMPTON STREET,
LONDON, W.C.2.

Dear Sirs,

Please supply ..... copy/copies of the current issue of EDUCATIONAL RESEARCH. I enclose cheque/money order/postal order for £..... .....s. .....d. at the rate of 5/6d. post free per copy.

Please also send me back issues as available*/future issues when ready. The remittance enclosed allows for.....issues at 5/6d. per issue post free.

This is a new/renewal subscription.

NAME.....

ADDRESS ..... BLOCK

..... CAPITALS

..... PLEASE

.....
Signature

To save unnecessary postage it will be appreciated if the remittance is sent with the Order Form. Cheques, money orders and postal orders should be crossed |& Co.| and made payable to Newnes Educational Publishing Co., Ltd.

* We regret that Volume One, numbers 1 and 2 are entirely out of print, and no further copies are available.

VOLUME IV, No. 1 November, 1961

EDUCATIONAL RESEARCH

A Review for
Teachers and all concerned with
Progress in Education

Published for
THE NATIONAL FOUNDATION FOR EDUCATIONAL
RESEARCH IN ENGLAND AND WALES

by
NEWNES EDUCATIONAL PUBLISHING CO. LTD.
Tower House, Southampton Street, London, W.C.2

THE MEASUREMENT OF PERSONALITY—I.

by F. W. WARBURTON

Department of Education, University of Manchester

Introduction

PHILIP VERNON has written in his *Assessment of Personality* that if we could but measure even three or four of the chief emotional-social and character dimensions and half a dozen or so of the chief interests and social attitudes we would probably be able to make as accurate predictions about personality as we already can about abilities by means of a short battery of ability tests. It is, of course, a commonplace observation among teachers that scholastic attainment depends as much on a pupil's approach to school subjects as on his innate capacity; in short, temperament, interest, personality, character, attitude and motivation, as well as ability, have to be taken into account in assessing a child's future progress in school. Unfortunately, although there is a growing volume of work on personality assessment, there is a striking discrepancy between the contribution that educational psychologists have made to this field and to that of the measurement of ability.

The study of personality is, of course, as old as human curiosity, and it has been the main concern of literature and the arts throughout the ages. From Greek times philosophers have attempted to systematise temperamental differences, and the Hippocratic fourfold division into sanguine, melancholic, choleric and phlegmatic still colours thought and language. During the present century, following the work of Freud, Jung, Adler and others, problems of personality maladjustment have largely become the province of psychiatrists concerned with the treatment of the mentally ill. The development of this branch of psychology has been haphazard rather than scientific; it has been a rich field for quackery, intuition, and global insights, as well as for men of the genius of Freud, whose psychoanalytic concepts have made a major contribution to world thought. But the work of the psychiatrist is largely conditioned by individual cases and circumstances. He has to take into account such matters as Mr. Jones' Saturday nights and Sunday mornings and his reactions to the price of eggs. The scientific control of social situations and behaviour is neither possible nor desirable when the main purpose is the alleviation of human suffering.

Advance along systematic lines has to be left to psychologists with facilities and time for research, and the opportunity to work with properly controlled samples of persons and personality traits. There are many different approaches to personality study within psychology. For example, Holland and Lindzey, in their book *Theories of Personality*, published in 1957, include chapters on Freud's psychoanalytic theory; Jung's analytic theory; the social psychology theories of Adler, Fromm, Horney and Sullivan; Murray's Personology; Lewin's Field Theory; Allport's psychology of the individual; organismic theory; Sheldon's constitutional psychology; stimulus-response theory; and Murphy's biosocial theory, in addition to the factorial theories on which the present article is based.

All these approaches have played an important part in recent work by psychologists on the nature of personality. The statistical method of attack described in the present article is only one among a dozen or so possible approaches to the subject, although in the opinion of the author, it is likely to prove the most fruitful in the long run.

Let us first define the term personality as it is used in personality *measurement* as distinct from the general psychological study of personality. Personality is usually held by psychologists to refer to the whole person. A common type of definition begins with the words, ' Personality is the sum total of,' and then proceeds to a list of innate dispositions, impulses, appetites, instincts, tendencies, habits, etc. Prince, for example, says that " personality is the sum of all biological innate dispositions, impulses, tendencies, appetites and instincts of the individual and the acquired dispositions and tendencies acquired by experience." Allport defines it as, " The dynamic organisation within the individual of those psychophysical systems that determine his unique adjustments to his environment." The Gestalt psychologists prefer this type of definition and object to the statement that an individual is merely an aggregation of parts, since this fails to take into account the integration of the complete personality. They regard a person as a ' whole,' but other psychologists have taken the view that science is necessarily analytical. Guilford has pointed out that, " It is humanly impossible for us to compare one ' person-as-a-whole ' with another ' person-as-a-whole '. The act of observation of a single person is an analytical process, as is the act of observing anything. Things, including persons, are known by their properties. An object is round, or sharp, or hard, or all of these things. A person is observed to react promptly, or vigorously, or accurately or in all of these ways." Certainly, for purposes of personality measurement, the term ' personality ' is usually taken to refer to the sum total of a number of discrete social and emotional qualities such as stability, liveliness, perseverance, carefulness and sociability, i.e., to traits based on the *way* in which an individual behaves in his social environment.

Scope of the present articles

This series of articles will be concerned only with the measurement of personality variables or traits, i.e., with certain social and emotional qualities that have been isolated and assessed by quantitative methods of analysis. The distinction between these characteristics and attitudes, interests and values, is rather ill-defined, but a social or vocational or educational psychologist would nevertheless have no difficulty in distinguishing between scales drawn up to assess personality and other characteristics. There is a common usage of terms such as interest and attitude to describe a given set of items, although different psychologists prefer to draw up different formal definitions.

Further, this series of articles attempts to deal with the relationship between measured personality and school attainment rather than with the emotional development of individual children. Both aspects are of great importance, but the study of the emotional development of children covers a wider field than personality tests, questionnaires and ratings, and requires a detailed and expert examination of the circumstances obtaining in each individual case. It is unwise for anyone other than a trained child psychologist or psychiatrist to attempt to ' analyse ' or treat children. A rigorous examination of the child's whole social background is essential. On the other hand, the school teacher has

an expert and first-hand knowledge of children's scholastic attainments and capacities, and he can help a great deal in investigations into the relationship between attainment and personality. Teachers have traditionally had to take children's attitudes to the school, to classwork, to other children and to the teacher into account, and it is clearly desirable that they should take an active part in research in this field.

Lastly, we have concentrated largely, although not entirely, on the British educational system, without extended reference to results obtained in America, since much of the schooling given in this country is radically different from that in the United States.

The general development of personality testing

The general strategy in research work on the assessment of personality has followed lines similar to those adopted in investigations into the nature of ability. The first task is to make a thorough survey of this area of psychology by examining the whole range of personality variables; the second, to establish the major dimensions (in the statistical sense) by means of factor analysis (correlational studies), i.e., to see which personality traits go together in the individual to make up the major factors; and the third, to attempt validation of the results by comparing them with the behaviour of people in everyday life. The largest contributions to personality assessment have been made by Eysenck, in Britain, and by Guilford and Cattell in the U.S.A. Cattell has carried out the most ambitious programme and in order to place recent work on the measurement of personality in British children in its proper setting, a very brief summary of his work is given below.

To begin his study of the personality field, Cattell carried out comprehensive surveys of previous research and literature. This is described in his book *The Description and Measurement of Personality* and the reader will best appreciate the huge scope of this work by reading it himself. Allport and Odbert have listed some 4,500 terms in the English language describing mental traits. By bringing together terms which the average user of the language would consider synonyms, Cattell found that this list could be reduced to about 160 terms. These formed a kind of basic English for the complete description of the personality. He next carried out in turn researches in the three main media for testing the personality, namely:

(1) ratings of people's behaviour in everyday life, made by other persons who are well acquainted with them, e.g., by fellow students in a University Hall of Residence;

(2) questionnaires concerning emotional and social reactions: e.g., Have you ever been afraid of the dark? Do you look forward to going to parties?

(3) objective tests, usually carried out under laboratory conditions, in which a person is set a task involving the personality, e.g., memorising under stress, or playing poker.

Ratings of everyday behaviour are, in some respects, the most important of these methods of assessment, as the ultimate criterion of personality differences is how people behave in society, not how they claim to behave according to a questionnaire, or how they

perform in standard tasks in a psychological laboratory. Cattell, therefore, turned his first attention to ratings, by searching the literature for factor analytic studies of the personality. For ratings there were 22 studies available (one of them by Cattell) involving 4,000 subjects and the extraction of a total of 88 factors. In Cattell's own investigation the subjects were 240 undergraduate women students in an American state university, living in sororities and independent houses. They were rated by their peers on 35 basic personality traits and 10 ' factors ' were extracted, as follows: Neuroticism, Cyclothymia, Perseverance, Dominance, Surgency, Intellect, Extraversion, Sophistication, Bohemianism and Excitability. [Factor analysis is a statistical method used for revealing the major influences underlining large numbers of observations, such as test scores. It was employed in this country by Sir Cyril Burt, Sir Godfrey Thomson, Professor Vernon and others to analyse the mental abilities.]

Cattell's work on ratings published in 1948 was an attempt to establish the major personality dimensions in the same way that Thurstone in 1938 had analysed the primary mental abilities. But it is obvious that future research cannot be confined to ratings, since, first, it is a highly laborious and inconvenient method. Secondly, it is not usually possible to find enough observers who are sufficiently acquainted with the subjects to make reliable judgements. Thirdly, the social environment is complex and cannot be adequately controlled. No two persons have precisely the same environment; students living in the same hostel, taking the same course of study nevertheless lead quite separate lives with a completely different sequence of experiences, and contacts with other persons. Fourthly, we clearly need a method of attack using ratings as a guide, but capable of more rigorous psychological and statistical analysis, and objective in the sense that all psychologists award the same scores for the same responses. At a later stage in research we thus have to study the personality under conditions which can be systematically varied and brought under experimental control.

Cattell, therefore, next examined 19 factor analytic studies of personality questionnaires, involving 4,800 subjects. There appeared to be 22 major personality factors in this material. Questionnaires suffer from certain disadvantages, of which the chief is that subjects can sometimes (but not always) fake their answers in order to give the psychologist whatever impression they wish about themselves, or they may be very unreliable judges of their own characteristics and deceive themselves easily. A third stage in personality research following the study of ratings and questionnaires is, therefore, the construction of objective, i.e., unfakable tests, which can be administered under standardised conditions in the same way as ordinary mental tests, but which nevertheless measure the major personality traits that characterise people in everyday life. About 60 factor analytic studies (five of them by Cattell), based on scores for over 8,700 subjects were taken into consideration. Twenty-three factors similar to those obtained for ratings and questionnaires were obtained from the examination of these data.

This work, which was published in 1946 (and which resembles in scope Spearman's somewhat less systematic survey of the cognitive field in his book *The Abilities of Man* of 1927) is based on the examination and collation of the findings from over 100 factor analytic investigations and the results for about 22,500 subjects. Cattell first tentatively hypothesised some twenty-five major personality factors, but the number has sub-

sequently been reduced to fifteen, on the basis of more recent research first published on ratings in 1947, questionnaires in 1950 and objective tests in 1951.

Rating Techniques

The development of personality measurement in this country is perhaps best appreciated by first considering in order, ratings, questionnaires and tests, and finally by assessing the validity and practicability of these measures for the teacher and the educational administrator. The present section is concerned with rating methods. These suffer from several drawbacks. When rating others, we are, in fact, recording person A's reactions to person B's behaviour, with double opportunities for error. Unless stringent precautions are taken, ratings are of little practical value.

The following steps should be taken in order to secure reliable and valid ratings.

(1) *Length of observation*

The rater should live with the subject, e.g., in a college, or be in constant social contact with him, and be able to observe him for a long time so that he is seen in many roles and situations. The period of observation should be at least two or three months (preferably a year), so that the rater can observe the subject in a number of unusual emotional situations. There is some evidence that rating reliabilities may go down as time is increased from, say, a few hours to several days, but reliability will rise again as the length of acquaintance is extended to periods of months and years. Persons seen briefly, as in an interview, usually assume a special role. Typically, the next fifty to a hundred hours of observation are spent in revising first impressions, and during this period the reliability drops, but high reliabilities of the order of ·90 are sometimes found with judgements based on long periods of time.

(2) *Definition of traits*

The personality trait to be assessed, such as ' perseverance,' should be clearly defined and exemplified by a list of ' critical incidents.' The judges, in short should have training in recognising, distinguishing and uniformly recording the traits they are asked to rate.

(3) *Use of normalised scores*

In order to avoid errors arising from differences in scatter among the ratings of different judges, it is advisable to express the ratings as ranks, and then to convert these to normalised scores by means of percentile tables, since a normal distribution can usually be assumed to be the most probable distribution of a complex, multiple-determined trait. This method equalises the mean and scatter for the scores of different judges, but not for different individuals or traits.

(4) *Number of judges*

Each subject should be rated by at least four, and preferably by twenty judges, whose assessments are averaged. The reliability of ratings increases rapidly with increase in the number of raters, for the obvious reason that individual prejudices and lack of

knowledge due to special personal relations with the subjects tend to cancel out. (It has been shown that statistically each new rater acts as an additional item when lengthening a test, i.e., the increased reliability follows the Spearman-Brown prophecy formula.)

(5) *Stereotypes and Haloes*

It is desirable to reduce stereotype and halo effects (giving a dog a bad or good name, and marking the same person up or down on several characteristics):

(*a*) by having, in some traits, the ' desirable ' pole, e.g., stability, rated high (e.g., ' 5,' with ' instability ' ' 1,') and in others the ' undesirable ' trait, e.g., neuroticism, rated high;

(*b*) by having *traits rated one at a time*, thus centring attention on the behaviour rather. than on the individual. One of the most effective methods of minimising halo effects is to make the assessments on one characteristic for the whole class of children, using slips of card on which are written the children's names and sorting these into the five categories, comparing the groups as they are formed. This procedure is far less conducive to ' halo effect ' than the assessment of one child at a time on a number of characteristics.

(*c*) Murray has pointed out a ' contrast ' error in ratings. This is a tendency for a rater who is himself toward one extreme of a trait to rate others more toward the other extreme. For example, a person who is himself extremely orderly and meticulous may tend to see and to rate others as lacking in these qualities.

(*d*) Biases may also exist in the opposite direction, that is, there may be a ' similarity ' error in rating. We often expect to find that others are like us in many respects. Landis found that tall raters tended to overrate the height of others, and emotionally unstable raters tended to overrate instability in others. The old saying that ' misery loves company ' perhaps applies in cases such as the latter.

This recommendation to assess one trait at a time throughout the groups is given in the N.F.E.R. handbook and also by many L.E.A.'s, but in spite of this, a very large proportion (roughly 90%) of teachers who were interviewed in connection with an N.F.E.R. survey by Walker said that they never used this method. A large number admitted that they had never even tried it and one thought that it was " just another psychologist's fad " (is 32° freezing point on a thermometer " just another physicist's fad? "). Some teachers said, " I prefer to consider the individual child as a whole," illustrating the confusion of thought which mixes up integrative methods of assessment suitable for clinical or diagnostic purposes and psychometric measures which require the isolation of single characteristics for statistical treatment.

(*e*) It is best if a day or two elapses between single-trait ratings. This is highly important, and a necessary precaution that is frequently

ignored by amateurs, with the result that they tend to give the same child a good or bad name several times in succession.

(6) *Independence of judgement*

The raters should be discouraged from discussing the subjects, since independent evaluation generally leads to considerably higher validity and reliability than team decisions, which tend to reflect the judgement of the team's more dominant members. Teachers making ratings of their pupils' behaviour in school can usually meet the above six requirements, except in respect of the number of judges. On the whole, it is best to confine the experiment to observers who know the subjects well, since two (say) teachers who are well acquainted with their pupils are more likely to make valid assessments, i.e., good predictions of their future behaviour, working by themselves than if they add the ratings of a third teacher who does not know the children well.

Ratings can take several forms. We can attempt simply to give numerical grades, say 1 to 5, or a percentage, or a mark out of 10 or make a mark on a graph and so on. Dr. Johnson is said to have been pleased at being assessed 20 out of 20 for Morality, although he was chagrined at receiving 0 out of 20 for Good Temper and Manners. Perhaps the two grades were not completely independent. Some modern referees even consider that their pupils have *no personality*. Psychologists rarely attempt to use more than seven or so grades, and five is probably the highest number of categories that can be used with efficiency.

There are many pitfalls for the untrained rater. For one thing, he tends to avoid the lower grades, thus reducing the discrimination of the scale; an A — E scale, for example, in which there are no E's is in effect an A — D scale, just as most examiners never award an A + mark, leaving it for the first Shakespeare or Newton they examine thereby lessening the sensitivity of their marking scale for everyone else. It is thus better to force raters to use a relative rather than an absolute scale. In selection procedures at the War Office Selection Board, grades of A (top 10%), B (next 20%), C (middle 40%), D (next 20%) and E (bottom 10%) were used. This method has the advantages of increasing discrimination and of approaching the normal (bell-shaped) distribution to be expected for all complex human characteristics based on large samples of persons. Another way in which the difficulty of biased distribution may be largely overcome is to give verbal descriptions of the statistical range such as Grade I, definitely above average, Grade II, average, and Grade III, definitely below average.

Examples of rating scales

The most reliable rating scales for children used in Britain are perhaps those of: (1) Fleming; (2) Schonell; and (3) N.F.E.R. Fleming (1954) has devised a 5-grade scale for the use of teachers, each grade being subdivided into + and —, making 10 points in all. She suggests that the traits assessed should be persistence, concentration, initiative, sociability and response in class, with a distribution very roughly approximating to 2%, 23%, 50%, 23% and 2%. Foundation Record Card, described by Walker, provides for ratings on stability, persistence, conscientiousness, co-operation and self-confidence.

Schonell has a rating scale for temperament for which he gives the following instructions.

(1) Rate a pupil in each trait by making a cross on the line at the point which you consider most accurately signifies his characteristics in that trait.

(2) Consider each trait separately, disregarding all others.

(3) Carefully read the description and type situations relating to each trait as given below.

(4) Consider each pupil in relation to other children of the same age.

Test 1

Self-Confidence

Extremely confident, amounting almost to 'cocksureness'	Very confident of own powers. Self-reliant	Confident	Lacking in confidence. Timid	Extreme lack of confidence. Over-dependent. Declines responsibility

Description

Self-confidence

In the positive form, this is indicated by self-reliance, ability to face difficulties, a feeling of assurance, ability to carry on independently and a readiness to take responsibility. In games and physical activities it is shown by boldness and courage. In its negative state it takes the form of timidity, doubt, over-dependence and lack of assurance.

The child who has confidence tries to proceed with a minimum amount of guidance; the child lacking in confidence must be aided constantly. The former likes to see what he can construct or produce after instructions are clear; the latter requires spoon feeding and actual assistance with his projects or problems.

Type situations

(1) Is he afraid of the dark?

(2) Can he look after himself (in personal habits) and his own possessions, or must he have people always helping him?

(3) Does he travel in a tram or bus by himself? (only pertinent in the case of children over nine years of age).

(4) Does he talk freely to new visitors?

(5) Is he good at games? Can he swim?

(6) Does the pupil answer with ease and assurance during oral lessons?

(7) Is the pupil good at dramatisation and reading in front of the class?

(8) Does he tackle new tasks well or does he constantly ask his companions and the teacher?

' Graphic ' scales are often considered to be an improvement. In the scale below of Cattell's H.S.P.Q. (High School Personality Quiz) the rater marks the spot with an X.

A	Warm-hearted	Average	Aloof
C	Stable	Average	Unstable
D	Phlegmatic	Average	Excitable
E	Aggressive	Average	Submissive
F	Lively	Average	Glum
G	Persistent	Average	Casual
H	Extraverted	Average	Shy
I	Sensitive	Average	Tough
J	Co-operative	Average	Selfish
O	Anxious	Average	Self-confident
Q_2	Self-sufficient	Average	Dependent
Q_3	Self-controlled	Average	Undependable
Q_4	Tense	Average	Calm

Kay used verbal descriptions based on Cattell's traits as follows:

Introversion-Extraversion

The extravert reacts more readily to the outside world. His behaviour and thinking are habitually governed with reference to his companions and the people around him. In contrast to the aloofness of the introvert, he is warm and sociable in his relations with others. He is likely to be ' thick-skinned ' and outwardly confident. In his work he is inclined to be less serious and adopt a happy-go-lucky attitude. He is talkative in class, whereas the introvert is difficult to ' draw out.'

Anxiety-Stability

The anxious boy is intense, excitable and serious, particularly under stress. In class he is likely to be easily upset, moody when corrected and prone to cry. He is particularly likely to show signs of stress when confronted with a new subject or situation, and will probably be timid at games. On the other hand, the boy who is rated high for stability

is tough and placid, insensitive to criticism and unaffected by new, unfamiliar situations. He is stolid in his general approach to work and in his social relationships. The Haggerty-Olson-Wickman Behavior Rating Schedule is an attempt to assess maladjustment in children. It comprises a series of graphic scales for rating thirty-five common traits. By analysing the numbers of well- and poorly-adjusted children who received each particular rating, a table of maladjustment indices has been derived. This scale can be used by teachers without special training in psychology, and it has the advantage that it is not necessary to assess personality disorders but only some of the more acceptable social and emotional traits.

Another type of scale is represented by the Vineland Social Maturity Scale, which contains 117 items, such as:

Reaches for familiar persons .. (4 months)
Dries own hands .. .. .. (2½ years)

Each item has been proved to be typical of the average (American) person of the age indicated. Unfortunately, this scale can be applied only by trained examiner who obtains the required information in an interview with a parent or someone else who knows the subject well. He checks the items that apply and works out a Social Age and Quotient in the same manner as a Binet Mental Age and I.Q.

The use of ratings in schools

Ratings have been little used in this country. The tendency is for each research worker requiring rating to make up his own items and most of them tell us more about the researcher's personality than the subject's. The variables chosen are not based on large-scale research work, such as that carried out by Eysenck, Thurstone, Guilford and Cattell, and merely assess fragments of the personality. We do not yet know which traits are positively correlated with success in school, and it is premature for educational research workers to choose their own variables, and thus prejudge the results, according to traits that the investigator hopes are relevant on *a priori* grounds. More seriously perhaps, other important variables which have been well-established by previous research are set aside in favour of the experimenter's private insights. Intelligence testing was at this stage in the last century. Moreover, if each investigator invents his own picture of the human personality, the findings in different researches cannot be directly compared, verified or rejected. In consequence, our knowledge of the personality, in so far as it is derived from ratings is as scrappy as our knowledge of ability would be if psychologists made up their attainment scales according to their personal views on the nature of ability, and ignored the work of Burt, Thurstone, Vernon, Thomson and others. Sir Cyril Burt says that " the traits to be assessed should be, as far as possible, mental abilities or character traits recognised in psychological research, not the vague entities or abstract nouns of popular description, which often designate an obscure compound of several traits varying more or less independently of one another . . . " Complex psychological problems such as the relationships between personality and school attainment cannot be solved by common sense and everyday experience, although in the initial exploratory stages these have a vital role to play.

The result of large-scale surveys by Cattell, Eysenck and Guilford, covering a decade of intensive and comprehensive research, suggest that the two main personality dimensions in both children and adults are (1) *neuroticism-stability* and (2) *extraversion-introversion*. It would, therefore, seem advisable to base teachers' ratings largely on these two variables. At the primary school neuroticism is most likely to show as nervousness, a tendency to show off, to tell lies and show ' spoiled child ' behaviour, whilst extraversion is likely to appear as friendliness and a willingness to co-operate with the teacher. In the secondary school, whether grammar, technical or modern, these traits will also be related to academic achievement, stability appearing as perseverance, conscientiousness and effort, and introversion as a studious disposition and a liking for books and academic pursuits.

Now as psychological research has shown, neuroticism and extraversion are quite separate traits and different pupils may show them in quite different degrees. *A* may be both highly industrious and of a studious disposition, *B* may be neither, being fickle and non-academic, while *C* may be a sticker although not much suited for intellectual pursuits, and *D* may be just the opposite being bookish but thoroughly dilettante and unreliable. By using rating scales based on these two major traits that are in keeping with the main trends of personality research we are more likely to arrive at fruitful and practicable findings which are of value in the everyday work of the school, just as teachers who are cognizant of the overall structure of the abilities are more likely to assess their pupils' school work on a realistic basis.

It is true that the N.F.E.R. survey described by Walker has discovered that the two most common traits assessed in British schools are perseverance and sociability which correspond very roughly to stability and extraversion, which indicate the two most important personality traits. The tortuous process of common sense and experience is thus slowly stumbling up the asymptote of truth. However, as the N.F.E.R. report points out, at the normal age of selection for secondary education, perseverance is not always a very highly developed quality in many of the brightest children, and a high degree of sociability may be a disadvantage from the point of view of scholastic success. Is it right to assume that, of two children equally matched in intelligence, the more persevering will make the better grammar school candidate? Is there not a possibility that *at the age of* 11 a lack of perseverance may be a natural concomitant of a lively and curious mind? These are typical problems that confront the practising schoolmaster and they cannot be scientifically investigated unless we have standard tests of perseverance and sociability comparable to, say, Moray House Intelligence tests, the National Institute of Industrial Psychology's tests of mechanical and spatial ability, or the N.F.E.R. attainment tests.

Allport states that, " there are approximately 18,000 terms (chiefly adjectives) in the English language designating distinctive and personal forms of behavior " and he gives some 150 examples beginning with the letter ' a '. Compared with such a list, the collection of nearly a hundred nouns and some sixty adjectives, found in the N.F.E.R. survey of records kept by local education authorities in this country is not particularly impressive, but even this number of terms emphasises the practical difficulties of analysing and assessing qualities of behaviour satisfactorily.

The N.F.E.R. conducted an enquiry into the keeping of school records. Under the general heading 'personality' were grouped all those sections of the records which deal with information about the child himself, other than that specifically concerned with intellectual ability and attainment, health and physique, or environmental influences. This is the sense in which the term is generally used on school record cards, and in addition or alternatively, the following headings appear:

Personal Qualities

(1) Special Characteristics
(2) General Behaviour
(3) Temperament, or Temperamental Qualities
(4) Development
(5) Disposition and Characteristic Moods
(6) General Attitude
(7) Personal Habits.

The enquiry showed that the number of Local Education Authorities making no provision for recording information about personality are: Infants 65%, Junior 39%, Secondary 54%, i.e., roughly half the authorities made no provision at all at any stage for recording 'personality' information. If comments about the children's personalities are not preserved on the cumulative record, there is little likelihood that they will be preserved anywhere else, and if information about personality is considered to be important, then the absence of a cumulative record with a section for such assessments or notes is a real loss. At the 11+ age, Yates and Pidgeon found that about three-quarters (73%) of the authorities stated that they instruct their teachers to take into account such information as they possess concerning their pupils' home backgrounds and qualities of character. 20·5% instruct their teachers *not* to take these factors into account. The remainder issue no instructions on this point.

The N.F.E.R. sent a questionnaire to the Local Education Authorities concerning the possible uses of assessments of personal qualities. The main points from the replies were as follows:

(1) *Diagnosis*

In cases of backwardness or maladjustment, diagnosis of the cause might be considerably assisted by a knowledge of the child's behaviour and development over the preceding years. Information of this kind might show long-standing weakness in some particular kind of reaction, or might show a turning point which could be related to other, possibly otherwise unsuspected, causes. (The value of personality assessments for clinical use is severely limited in the opinion of the writer, but there is no reason why they should not give useful pointers to action, in the same way as any other external assessment. Objective tests, and possibly standardised questionnaires, might give very valuable information.)

(2) *Value to the teacher*

A knowledge of previous assessments of ' personality ' may help a teacher, as soon as the child comes to the class, to utilise quickly strong, useful characteristics, forestall the development of undesirable ones, and encourage desirable ones, which may be only weakly developed, without a period of delay for the personal discovery of such characteristics.

(3) *Prediction*

If a child's normal mode of behaviour has been consistent, a teacher may be in a position to predict the probable course of the child's development, and in particular make an estimate of his chances of success in some specific form of education. Marked variability of behaviour, or instability, may also lead to a prediction of ' trouble ahead ' unless remedial measures are instituted.

With regard to grammar school selection, it is generally agreed that the extent to which a child is likely to be successful in school depends not only on his level of ability or the standard of attainment that he has reached by the time he leaves his primary school but also on various personal qualities. His capacity for concentration and for persistent effort, his adjustment to the school environment, the degree of confidence that he possesses in his own ability, the breadth of his interests and his liking or disliking of school life are all important factors. Admittedly, as long as there is competition for grammar school places, there is a strong argument for allotting them to pupils who not only have ability, but who have already demonstrated that they know how to use it, and are willing to do so. This, however, is not a good reason for making the possession of certain character qualities a criterion for entry to grammar schools. It may be argued that the grammar school should at least have the chance (some would say the right) to train such qualities of character, when they are lacking, as will enable the child to use his abilities to the full. Moreover, it seems reasonable to stress character qualities in selection at the border line, where children are known to be of grammar school capacity, and where the written examination has failed to discriminate adequately between them. On the other hand, no child who shows the *requisite ability* should be denied the opportunity of taking a grammar school course on the grounds that he lacks certain of the personal qualities that are regarded as desirable or even essential, for academic success. It is possible that an alleged lack of persistence may be due to boredom with the restrictions of the primary school course, and he may be regarded as apathetic when, in fact, the stimulus provided by a new school might release unsuspected energy and enthusiasm. Unless the qualities that are assessed can be shown to be stable there would seem to be little justification for taking them into account.

Some objections to ratings

A common objection to using trait ratings for educational purposes, is that, although the assessments in terms of traits may classify and grade the children's behaviour, they give no indication of its causes, and may group together children who are basically very different, although they may show superficial resemblances. This criticism shows a

complete misunderstanding of the nature of scientific measurement. The study of the causes of children's behaviour is a separate problem. No two persons, for example, have identical mathematical ability, simply because they have the same examination marks, their eyesight is not the same because they have identical opticians' cards; no two rooms have the same atmospheric conditions because their barometric pressures are identical. Assessments of personal qualities, like any other quantitative measures used in science have to be accepted and validated as they stand. Their underlying structure is an important, but different, consideration, and a matter in psychology for the clinical psychologist. It is important if fifty introverts show a significant tendency to have higher academic attainment than fifty extraverts, even if the introversion has a different background in all fifty cases.

Another objection to ratings is that, although in some children at least, certain forms of behaviour have become characteristic and predictable by the time they reach school age, nevertheless, much of the childs behaviour is a reaction to immediate circumstances, and may be expected to vary according to social contacts with different children, teachers, or in different places and circumstances, e.g., home and school. This is no doubt true. Psychological qualities vary a great deal according to the situation in which a person finds himself, but on the other hand, they do not vary at random. Personal characteristics have some measure of consistence, as well as variability. There may well be circumstances in which General Montgomery would be more introverted than Clement Attlee, but they are probably few. It is safe to predict more withdrawn behaviour from Attlee than from Montgomery in the future, and it would be nonsense to predict the opposite, or to say that no such differences can be successfully predicted.

Another difficulty lies in the nature of the distribution of personal qualities. One local education authority (according to the N.F.E.R. survey) expects 5% of its junior school children, i.e., 2 or 3 in each class, to be ' totally unreliable, quite irresponsible; a cheat—faithless and deceitful; untruthful—an habitual liar; dishonest—guilty of petty thieving; frequently blames others for his own misdeeds; rarely carries out duties or observes rules '. Some of the characteristics assessed in school records are determined almost entirely by cultural or environmental influences, and in these circumstances it is not realistic to expect, say, only 30% of the population to be co-operative! It is the descriptive phrases used on school record forms which are misleading and distort the natural normal(ish) distribution, which is almost certain to be found for any complex psychological characteristic.

Research into school ratings

The N.F.E.R. inquiry into selection, conducted by Yates and Pidgeon, asked teachers for ratings of their pupils' persistence, sociability and self-confidence, choosing these three traits because they appear frequently in record cards used by local education authorities. Questionnaires were also devised for use with a borderline group in which typical manifestations of these qualities were described and the teachers were asked to indicate for each pupil whether they had observed this kind of behaviour frequently, occasionally or hardly at all. Neither the teachers' ratings of these qualities, nor the description of behaviour, yielded a satisfactory measure. Yates and Pidgeon came to the

conclusion that there are as yet no means available to education authorities of effectively identifying and assessing the personal qualities that influence a child's chances of success in a grammar school course. In fact, the N.F.E.R. recommends that an authority that deems it appropriate to include judgements of personal qualities in its procedure for allocation would be well advised to invite its teachers to make a global assessment of overall suitability for grammar, technical or secondary modern schools. The teachers with whom this problem was discussed were unanimous in the view that expert guidance was essential and that it should be conveyed by means of illustrations of concrete items of behaviour rather than by definitions in abstract terms.

In the N.F.E.R. enquiry on school records, it was found that in Junior schools, where it is possible to compare children of all levels of ability, the correlations between school work assessments and almost any ' personality ' quality was uniformly about ·6. That is, there appeared to be a fairly strong tendency for assessments of personal qualities to go with good ability in school work. These correlations are fairly high, but they are probably due largely to ' halo ' effect. The results were as follows:

Comparison of Persistence Rating with Second Year School Work

Persistence Rating	*N* (78)	*Percentage in Upper (School Work) Group*	*Percentage in Lower (School Work) Group*
12 and above . .	. . 50	80·0	20·0
11 and below . .	. . 28	50·0	50·0

It is clear that there is a substantial relationship between the personality measures and the criterion (the tetrachoric correlations are about +·60), but it must be remembered that these assessments may provide little *additional* prediction when added to a battery of more orthodox selection tests. Evidence on this point is essential. More convincing evidence would be significant correlations between objective tests of ability and attainment.

MacClelland describes a research in which teachers were asked to grade all pupils in regard to industriousness. He found correlations of +·598 (boys) and +·647 (girls) with I.Q., and +·568 (boys) and +·664 (girls) with attainment. The high correlation with I.Q. suggests considerable ' halo ' effects. The study also showed that the girls were more industrious than the boys, although the difference was not very great. The good boys were nearly as industrious as the good girls, but the boys of poorer ability were less industrious than girls of the same type. On the whole, if it were not for the regrettable fact that the observed relationship between personality assessments and school attainments may be largely due to halo effects, the overall correlations of the order of size of +·6 would be extremely satisfactory, and little inferior to those obtained by tests of ability. But since they add little *additional* prediction to estimates of attainment, they have little practical value. The results of attainment tests given at 11+, particularly those of the ' old-type ' and essays may well act as incidental measures of those personality traits which are relevant to success in the secondary school.

Thorpe and James investigated the relationship between teachers' ratings of adjustment and maladjustment, and maladjustment and more objective estimates obtained by psychological testing. Three groups of 36 children were studied—a group of the most adjusted, a group of the least adjusted and a normal group selected at random. Objective tests of neuroticism, which had been shown previously to distinguish between neurotic and normal children, were found, with the exception of the sociometric test (the children's liking for one another), not to differentiate between these three groups. It seemed likely, in fact, that the teachers were making their judgements on the basis of extraversion rather than neuroticism. Hallworth has shown, however, by analysis of the intercorrelations between another set of teachers' ratings that their judgements were based on two dimensions similar to Eysenck's and Cattell's two major factors. They can be named (1) ' extraversion ' or (2) ' reliability and conscientiousness ', i.e., stability. Kay found correlations between teachers' estimates and Cattell's personality factors of +·76 (Extraversion) and +·46 (Anxiety) in one school, but of only +·15 (Extraversion) and —·08 (Anxiety) in another school, in which the children were not so well known to the teacher. This suggests that only those teachers who know their pupils well should rate them; any attempt to increase the number of raters is unwise.

Bibliography

Works Cited in the Text (Part I)

General Personality Field

ALLPORT, G. W. (1937). *Personality: A Psychological Interpretation.* New York: Holt.

BURT, C. L. (1950). In REYMERT, M. L. (ed.), *Feelings and Emotions.* " The Mooseheart Symposium in co-operation with the University of Chicago." New York: McGraw Hill.

CATTELL, R. B. (1946). *The Description and Measurement of Personality.* New York: World Book Co.

CATTELL, R. B. (1947). " Conformation and classification of primary personality traits." *Psychometrika*, 12, 197-220.

CATTELL, R. B. (1950). *Personality: A Systematic Theoretical and Factual Study.* New York: McGraw Hill.

CATTELL, R. B. (1951). " A factorisation of tests of personality source traits." *Brit. J. Psychol. (Stat. Sect.)*, 4, 165-178.

HALLWORTH, H. J. (1960a). " Teachers' Ratings of Personality among Secondary School Children." Paper read to Int. Congr. Psychol., 1960.

HALLWORTH, H. J. (1960b). " Anxiety in Secondary Modern and Grammar School Children." Paper read to the British Psychological Society.

KAY, M. (1960). " A comparison between the personality test scores of secondary modern children in a Jewish and a non-Jewish school." Diploma in Educational Psychology dissertation, Department of Education, University of Manchester.

LANDIS, C. (1930). " Questionnaires and the Study of Personality." *J. Nerv. Ment. Dis.*, 83, 125-134.

MCCLELLAND (1949). *Selection for University Education.* London: University of London Press.

MURRAY, H. A., *et al.* (1938). *Exploration in Personality.* New York: Oxford University Press.

PRINCE, M. (1924). *The Unconscious.* New York: Longmans Green.

SCHONELL, F. J. (1944). *Backwardness in the Backward Subjects.* Edinburgh: Oliver and Boyd.

SPEARMAN, C. (1927). *The Abilities of Man.* London: Macmillan.

THORPE, J. C. and JAMES, D. P. (1957). " Neuroticism in children, I. An investigation of normal and neurotic group differences." *Brit. J. Psychol.*, 48, 26-34.

THORPE, J. C. and JAMES, D. C. (1960). " Neuroticism in children, II. Identification and measurement of a neuroticism factor." *Brit. J. Psychol.*, 51, 153-157.

THURSTONE, L. L. (1938). *Primary Mental Abilities.* Chicago: University of Chicago Press.

VERNON, P. E. (1952). *Personality Tests and Assessment.* London: Methuen.

WALKER, A. (1955). " Pupils' School Records: A Survey of the Nature and Use of Cumulative School Records in England and Wales." *N.F.E.R. Publications No.* 8. London: Newnes.

YATES, A. and PIDGEON, D. (1957). " Admission to Grammar Schools." *N.F.E.R. Publications No.* 10. London: Newnes.

CRIME, CAUSE AND TREATMENT: RECENT RESEARCH AND THEORY

by LESLIE T. WILKINS

Mr. Wilkins is a member of the Home Office Research Unit, but the views expressed here are his own.

MOST people have the impression that crime in general and juvenile delinquency in particular has increased rapidly during the last few years and is today a very serious problem indeed. Is this impression correct? If so, what can be done about it? Certainly the number of 'crimes known to the police' was, indeed, in 1959 (according to the most recent figures available, given in *Criminal Statistics, England and Wales*[1]) the largest ever recorded in our history.

But as Huff points out in his admirable textbook *How to Lie with Statistics*[2] "averages and relationships, trends and graphs are not always what they seem. There may be more than meets the eye or there may be a great deal less." Huff records that the official figures for encephalitis in the Central Valley in California in 1952 showed triple the figure for the worst previous year. Many alarmed residents "shipped their children away. But when the reckoning was in, there had been no increase in deaths from sleeping sickness. What had happened was that the state and federal health people had come in in great numbers to tackle a long standing problem, and as a result of their efforts a great many low grade cases were recorded that in other years would have been overlooked, possibly not even recognised." Can there be a similar phenomenon underlying crime statistics?

In considering this, it is useful to examine rather more closely the background of the figures for crime. It is, of course, immediately apparent that the police do not become aware of crimes by some mysterious process of 'knowing'; nor do they get to know about them by direct observation if the criminals are able to avoid it. Crimes 'known to the police' are usually reported to them by the public, most frequently by the victim. This means that an event must first of all be *defined by a member of the public* with no special legal knowledge, as a crime, *and* as a crime such as to justify his telling the police about it. Now the legal definitions of 'crimes'* are clear enough but extremely wide in scope. For example, the value of a stolen article is not relevant to the legal definition of larceny; robbery with violence may involve nothing that the lay citizen would classify as violence at one end of the extreme limit of the definition, while at the other end the worst possible headline story would fit the same legal category.

Few children have not been fortunate at some time in finding some small coin or other apparently discarded article which they have retained and in consequence are guilty of larceny by finding. On the other hand, vandals could do very considerable damage to your car, but the incident would not be recorded as a 'crime'* unless the damage done amounted to more than £20; but if instead of damaging your car they merely opened the door and removed a small packet of cigarettes from the glove pocket, they

*For purposes of 'crimes known' only indictable offences are considered.

would have committed a crime of larceny from a vehicle. Legal definitions of *crime* are intended for use in legal processes and are doubtless adequate for those purposes, but they are not directly translatable into sociological indices. Included in what is usually regarded as normal behaviour are some acts which are crimes by legal definition, but at the extreme end of a distribution are crimes which all sane persons, even offenders themselves regard as such.

Crime is a human action, and it is convenient to try to set up a verbal or mathematical model of actions which will help in discussions of the question of whether crime has or has not increased. It is possible to picture human actions as forming a frequency distribution, a continuum from very saintly acts to the most sinful. There are very few extremely saintly acts committed in the world, and there are very few extremely sinful acts, the majority of our actions are just simply normal. In fact it does not seem unreasonable to postulate that if we could measure the degree of saintliness or sinfulness of actions, the ethical content would be normally distributed in much the same way as we arrange intelligence measurements to be normally distributed.

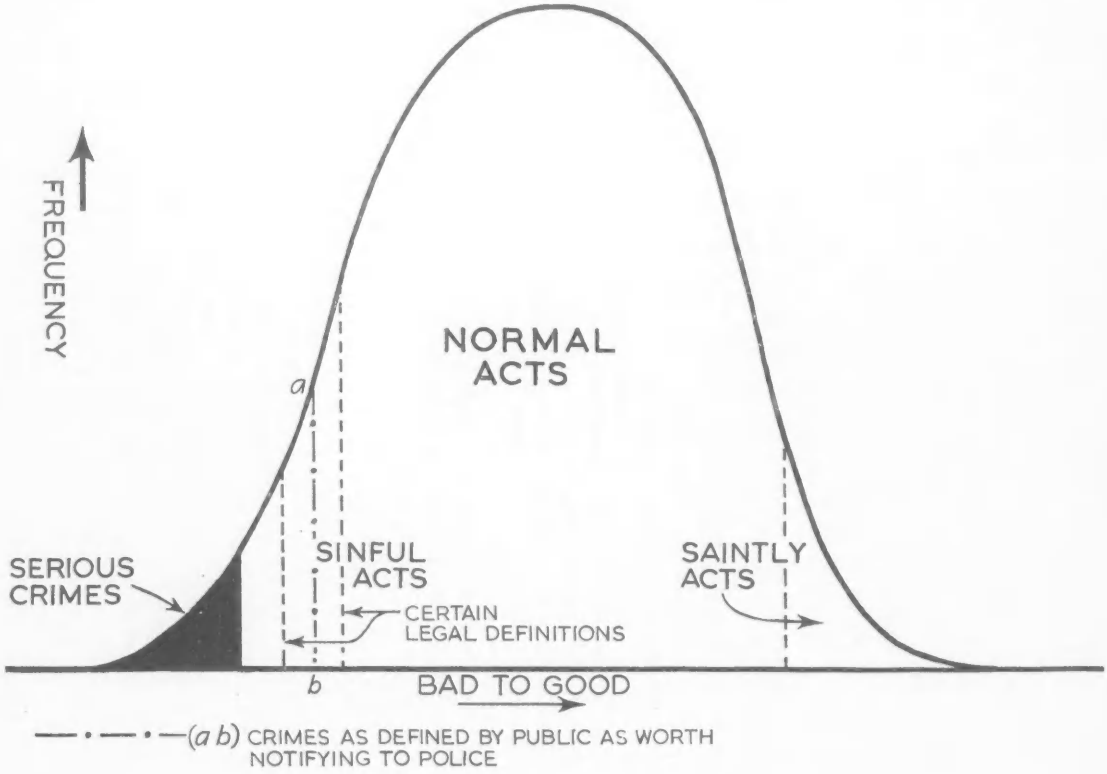
A diagram of the familiar distribution is given as figure I.

If we now take the mathematical model afforded by the normal frequency distribution shown as figure I, for purposes of argument, to represent the distribution of the ethical content of human actions, crimes will be represented, in the main by actions at the ' sinful ' end of the scale. But the point of cut-off between crime and no crime is of importance in counts of the frequency of notification of crimes known to the police. At this stage it is possible to introduce another ethical consideration. It is often represented that the main ' cause ' of the current ' crime wave ' is the drop in moral values of present day society. No less an authority than Lord Pakenham has expressed such a view. If this argument is not circular, is it a fact that today the average citizen is more likely than in any previous years to become the victim of a criminal attack upon his property or person?

It is known that ' crimes known to the police ' do not include a very large proportion of all the crimes committed. This is because the figures upon which the ' crimes known ' figure is based are derived from the sample of crime which has been reported. That is to say the incident before being included in the ' crimes known ' figure, has to be *both* socially defined as worthy of notification, and defined by the police as a crime in the technical meaning of the term. It would be expected, for example, that quite a few incidents of vandalism are defined by the public as worthy of notification, but are not in fact crimes and do not find their way into the statistics. On the other hand there are many thousands of incidents of shop-lifting which are certainly ' crimes ' but are defined by the victims as ' not worth reporting.' There is, too, considerable scope for social definitions to vary both from place to place and from time to time without any change in the underlying situation.

Any change in the social definition of an event worth notifying to the police will, in completely stable conditions, influence the bias of the sample of crimes represented by the ' crimes known ' figure. It is possible to argue that if our moral standards are improving—that is to say if we are becoming *more* stringent and requiring a *more* conforming standard of behaviour—the definition of the social concept of a crime that should be

Figure I


FREQUENCY
NORMAL ACTS
a
SERIOUS CRIMES
SINFUL ACTS
SAINTLY ACTS
CERTAIN LEGAL DEFINITIONS
b
BAD TO GOOD
(a b) CRIMES AS DEFINED BY PUBLIC AS WORTH NOTIFYING TO POLICE

notified to the police will tend more closely to agree with the legal definition so far as the majority of crimes are concerned, since by far the biggest component in the ' crimes known ' figure is that of larceny. In terms of the model of figure I, this would mean that the line *ab*, would move further towards the right so that more of the area under the curve would be taken in. The shape and the size of the curve need not change.

As Lodge[3] remarks " only large changes or persistent trends can be relied on with any degree of safety to have a cause which it is worth while to investigate. There should be borne in mind the possibility of factors such as improved police methods, greater readiness of the public to report the offence, a court decision affecting the name given to an offence, or other such changes." While statistics of crimes are very difficult to interpret, statistics relating to offenders who appear before the courts are recorded with substantial accuracy. Taking a very broad view and using all the information available, it is only fair to say that it is likely that the crude figures of ' crimes known to the police ' which have been quoted in many quarters to indicate a substantial increase in crime rates of recent years, do in fact represent a real increase. It is even more certain that particular age groups of young persons are causing justifiable concern.

The Causes of Crime and the Treatment of Offenders

The Criminal Justice Act, 1948, Section 77 (i) b, gave the Secretary of State the authority to incur expenditure " in the conduct of research into the causes of delinquency and the treatment of offenders, and matters connected therewith." The wording of this section is interesting. It separates the concept of the " causes of crime " from the concept of the " treatment of offenders." It has already been noted that it is very difficult to discuss the meaning of ' crime ' and almost impossible to measure it. The problem of the causes of crime is a very large, perhaps unlimited, problem. The treatment of offenders although involving some extremely difficult concepts has at least the limitation that an offender is a person! The distinction between the two types of criminological research, that which is concerned with and limited to persons and that which is concerned with a broad and general and perhaps undefinable concept, is important. Let us first consider what is known or believed about the causation of crime.

It is perhaps surprising that with a subject about which the experts know so little, the layman knows so much. Ask any normal citizen about the causes of crime and it is certain that you will get a quantity of positive replies. Ask how offenders should be treated so as to ensure that they do not commit further offences and few will have any doubts about the methods they advocate, unless perhaps they have made a study of the subject! Degrees of certainty of belief in this field seems to be inversely proportional to knowledge of facts. A recent study by the Home Office Research Unit[4] led to the suggestion that children of about five years of age were more likely to be disturbed by war or social upheaval than children of any other age. The *Daily Express*[5] commented on this research. " Home Office Research workers labour on the problem of delinquency among young children. They put forward a great many figures and arrive at a conclusion which will surprise nobody . . . " Similarly *The Daily Mail*[6] commented, " Clever people sometimes go to enormous trouble to reach conclusions which simpler minds have already jumped to." *The Observer*[7], however, remarked that " one of the findings, that

disturbances around the age of four and five are particularly harmful seems of a doubtful validity." Thus, for some persons the conclusions were obvious without any evidence, and for others, despite the evidence, the conclusions were still untrue!

This state of affairs is typical of the whole field of criminology*. For every theory concerning the causes of delinquency, no matter how strongly held, it is possible to find a contradictory theory equally strongly held. Some persons have, therefore, proposed to do without theory and to ' stick to facts.' But it is not clear what action this might mean without the cement of a theory to hold the facts together. There are, of course, good theories and bad theories, and there are ways of distinguishing them. A good theory can assist the processes of scientific works. Some theories of the causes of crime will be examined mainly to test whether they are good or bad theories—whether or not they fulfil the role of a theory—not whether they are true or valid.

How should we examine a theory? The best theories are concise, elegant summaries of observations; the observations are accepted as sound and they have been subject to replication by different workers. Theory interrelates observation in a common statement. The best way to distinguish good theories from bad is to consider ways of destroying them. In the first place we should examine the theory to see what deductions can be made from it to set up postulates which may be tested either experimentally or by observation of the facts of life. That is to say, the first step in testing a theory is to state the practical consequences of it. If the operational definitions (deduced practical consequences) are proved untrue the theory is discredited. The value of a theory is directly related to its relevance to the particular problem under study, and the consequences and operational definitions which can be deduced from it. No theory can ever be proved to be true; it is held for so long as no better theory can be found. Clearly the greater the number and the wider the scope of the hypotheses and practical consequences that can be deduced from the theory and the greater the number and the wider the scope of the hypotheses which have been or could be tested and which *could* be or *could have been* proved false, the better is the theory.

Two theories of delinquency

The first theory most vigorously put forward in current criminological literature is that of ' multiple causation.' This merely asserts that crime is a function of numerous factors. Examination of this ' theory ' by the criteria given above suggests that it is a statement that can hardly be dignified by the term ' theory.' Perhaps to some extent it summarises certain observations, although this is not clear. Proponents of the theory would claim that since research has found no single factor as the cause of crime, but rather many different factors appear together associated with greater or lesser frequency of criminal activity, this is evidence for multiple causation. But the theory does not facilitate the deduction of any hypotheses or practical consequences that are of any help at all. If it is claimed that the theory applies to all factors which are operationally found to be related to criminal tendencies as they become known, it is clear that the theory lacks the

*The analogy between research into the causes of delinquency and research into education is close—pontification is readily based on prejudice and awkward facts are denied or ignored.

major and essential feature of a scientific theory—it is framed in such a way that it is impossible to find any test whereby it could be proved wrong! The theory of multiple causation is, then, no theory. At best it could be an anti-theory which proposes that no theory can be formed regarding crime. Such a council of despair may be legitimately put forward as a kind of ' beatnik ' philosophy, but not as a scientific theory.

A theory which does meet the tests proposed above and which is attracting considerable attention at present is due to Sutherland[8] who states, " a person becomes delinquent because of an excess of definitions favourable to violation of law over definitions unfavourable to violation of law." This is called the principle of " differential association."

This means, in terms of very broad principles that it is believed that when persons become criminal they do so because of contacts with criminal patterns and also because of isolation from anti-criminal patterns; that persons assimilate the surrounding culture unless other patterns are in conflict. From this theory a number of hypotheses can be set up and subjected to test. A recent examination of this theory by Glaser[9] reveals that it is not only a good theory in that a large number of practical consequences (hypotheses) can be derived from it, but that a large number of these have been demonstrated by other research workers to hold up under test.

There are, however, a number of other factors known to apply to some criminals which it is not possible to deduce from the differential association theory. What then should happen to the theory? As James Conant recommends " a conceptual scheme is not to be discarded merely because a few stubborn facts cannot be reconciled; it should be modified, unless it can be replaced by a better one." Is it possible to find a better theory or to modify this one? This question may be examined by reference to some further theories and general conceptual frameworks.

Perhaps the best known work on the gang problem is that by Cohen in *The Culture of the Gang*[10]. The theory put forward in this work is summarised by Kitsuse and Dietrick[11] as follows:

(*i*) " The working-class boy faces a characteristic problem of adjustment which is qualitatively different from that of the middle-class boy.

(*ii*) " The working-class boy's problem is one of ' status frustration,' the basis of which is systematically generated by his early exposure to the working-class pattern of socialisation.

(*iii*) " The working-class boy's socialisation handicaps him from achievement in the middle-class status system.

(*iv*) " Nevertheless, he is thrust into this competitive system where achievement is judged by middle-class standards of behaviour and performance.

(*v*) " Ill-prepared and poorly motivated, the working-class boy is frustrated in his status aspirations by the agents of middle-class society.

(*vi*) " Delinquent sub-culture represents a ' solution ' to the working-class boy's problem for it enables him to ' break clean ' with the middle-class morality and legitimises hostility and aggression against the sources of his frustration.

(*vii*) " Thus the delinquent sub-culture is characterised by non-utilitarian malicious and negativistic values as ' an attack on the middle-class where egos are most vulnerable . . . It expresses contempt for a way of life by making its opposite a criterion of status.' "

Many writers have found difficulties in reconciling this theory with observed facts, even in the U.S.A. It is not possible to state all the objections here, but a modification of the theory is interesting in that it relates in obvious ways with the theory of differential association to which reference has been made earlier. This theory is due to Kitsuse and Dietrick who were concerned with the difficulty in Cohen's theory of explaining the continuance of the delinquent sub-culture. They postulate:

(*i*) " The individual learns the values of the delinquent sub-culture through his participation in gangs which embody that sub-culture.

(*ii*) " The motivation of individuals for participation in such gangs is varied.

(*iii*) " The malicious non-utilitarian and negativistic behaviour which is learned through participation in the sub-culture is met by formal negative sanctions, rejection and limitation of access to prestigeful status within the middle-class system.

(*iv*) " Thus, participation in the delinquent sub-culture creates similar problems for all its participants.

(*v*) " The participants' response to the barriers raised to exclude them from status in the middle-class system (that is the ' problem ') is a hostile rejection of the standards of respectability and an emphasis upon status *within* the delinquent gang.

(*vi*) " The hostile rejection response reinforces the malicious non-utilitarian and negativistic norms of the sub-culture."

Readers may find it interesting to compare the two theories stated above with those of Bernstein in *Educational Research*, Vol. III, No. 3. These theories discuss the learning of values, but do not consider the mechanics of learning. As Bernstein aptly states, the problem is " how the outside gets into the inside." If we can find out something about this, we may be able to intercept the process, analyse it effectively and then find something that can be done about it. Bernstein suggests that language is the main medium. It would appear that there is some evidence in favour of communication factors in some aspects of delinquency in this country. But before they can be considered it is necessary to look at one further theory.

The theories of delinquency which are currently attracting most attention are based on sociological concepts. There are theories which are basically psychological. These are well known, particularly in this country those associated with the work of Bowlby, Andry & Stott. In the U.S.A. sociology and psychology are closer together in terms of their conceptual models and the social psychologists have adapted sociological theories of delinquency to psychological frames of reference. Theories of perceptual processes and particularly the concepts of social perception have particular promise. This is an interesting development and should, perhaps, be noted and illustrated.

It is assumed that all ' knowledge ' is acquired indirectly; man constructs his knowledge by orderly transformation of coded signals, and this process is called cognition. Intuitive directness, whether in perceiving or thinking may result from practice and predisposition, but it is achieved by inferential processes, no matter how short-cut they may become through learning and development. The attempt to identify an approaching figure, to develop an idea, to form a sentence or any other activity, involves selecting or constructing a means appropriate to some end. Our perception is conditioned by experience. Social perception refers to the way in which the world of other people is seen—how people try to make sense of others. Social perception is the process of observing, recording (coding and storing) and organising the experiences one has with other people, groups and social organisations. The term is also used—and this is important—to refer to the way social factors influence observations of one's world. Work in this area is concerned with such matters as how people ' understand ' other people and how they relate this ' understanding ' to adjustments in their own behaviour, how accurately people observe, describe and judge other people and themselves. Differences in ability to perceive are often related to ideas of ' social maturity.'

One theory of delinquency which has been transformed into a perceptual ' model ' is that of ' anomie.' This theory is related to the work of Cohen discussed earlier. The concept of ' anomie ' was first put forward by Durkheim[12] to explain variations in suicide rates. It was a sociological theory but it may be easier to understand and test in its transformed style. The concept is stated by Srole[13] in perceptual terms as follows:

A person having a high ' anomie score ' would perceive:

(*a*) the lot of the average man as getting worse;

(*b*) the future so black that it is hardly fair to bring children into the world;

(*c*) public officials as uninterested in problems of the average man;

(*d*) other people as untrustworthy;

(*e*) life to be lived very much for today—tomorrow to take care of itself.

High ' anomie ' has been related to the criminal gang phenomena of New York and other large American cities. It is argued that ' anomie ' results when an individual is prevented from achieving his objectives, and is in this way related to the ' blocked opportunity ' explanation of the gang. Perhaps in the U.S.A. where all levels of society are attracted by the ' American dream,' where the values sought by all are the middle-class values, ' blocked opportunities ' may be a more valid theory than in this country where most sociologists have reported not excess of ambition among young people, but a disturbing lack of it. Or is, perhaps, the ' American dream ' not universal even in the U.S.A., and is this merely another method whereby the middle-class social workers and researchers project their own value structure to others?

Perceptual process theories are capable of high degrees of generality. For example Fiedler and Kahn, Mannheim and Rabin of the Group Effectiveness Research Laboratory at Illinois[14] have studied particularly ' perception of others,' ' reference groups and membership groups ' and the ' self-image ' and found these to be valuable concepts in relation to working groups in the armed forces, church groups and even groups in prison.

The theories of perceptual processes are related to problems of communication. Perception involves transformation of coded signals (sight, sound) and implies an apparatus analogous with the radio or television receiver. Communication is concerned with both transmission and reception. Market research workers are making very effective use of these theories by exploring ' brand images ' (how brands are perceived) and the use of symbols having various emotional content.

Symbols, images, communication and perceptual processes are likely to have some meaning in exploring delinquent and deviant behaviour. For example, group and individual stress or tension impedes communication and perception. Glaser has recently suggested that the differential association theory would be better if modified to differential identification or differential expectation. While the differential association theory stresses the factor of learning, differential identification allows not only for learning but for modifications of learning through perceptual processes interacting with personality differences.

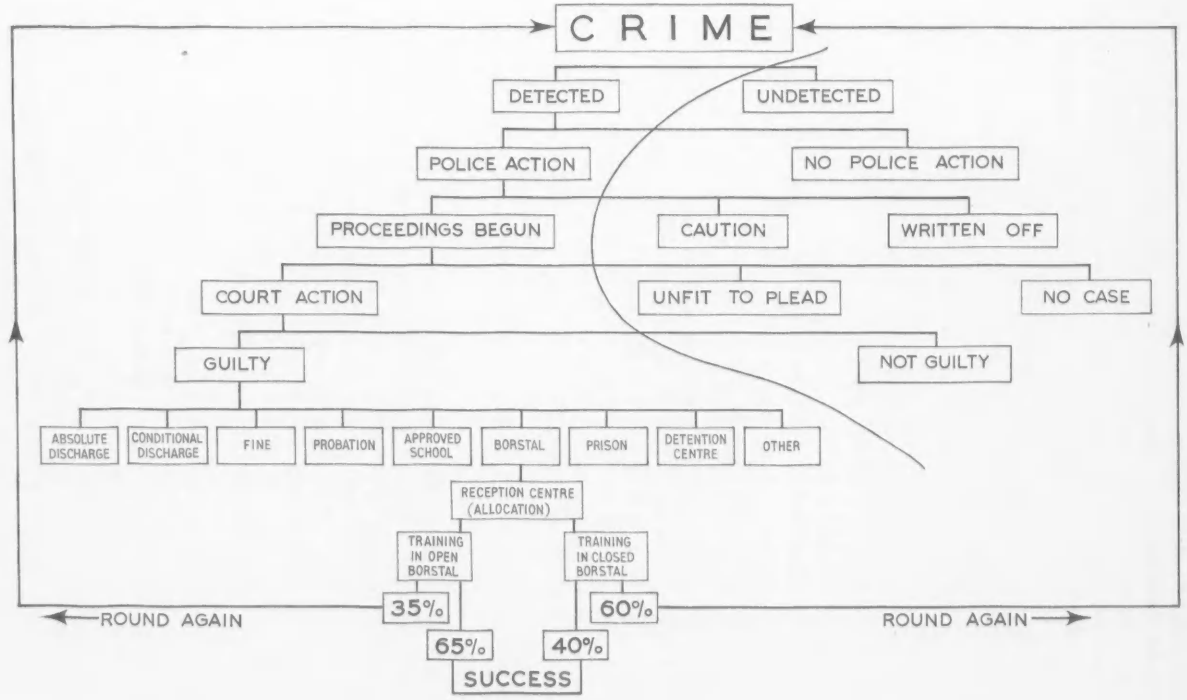
In the Home Office Research Unit report, mentioned briefly earlier, it was shown that youths of between 17 and 21 in 1955 and onwards, who would have been expected because of their years of birth (' delinquent generation effect ') to be exceptionally delinquent, were, in fact even more delinquent than could be estimated by the year of birth analysis. Some other factor peculiarly conducive to delinquency among young men has apparently been operating during the last few years. Although an accretion of effect is another explanation, the most direct inference that can be drawn is that this ' other factor ' had something to do with the fact that about 1955 the young adult males in the particular sub-cultures which give rise to most delinquency, hit upon a ' communication symbol '—a particular form of dress (jeans, drapes and so on). As with any other minority group, symbols of communication are powerful factors in preserving the group. There is no doubt, too, that the easy identification which the dress symbol made possible, both to persons inside and outside the minority group, made the process described above as " formal negative sanctions, rejection and limitation of access " easier to apply by persons outside the group and increased the identification and solidarity within the group. Thus, delinquents became more delinquent.

Theories are attractive, but they are and should be highly vulnerable. The present writer has no wish to suggest that, because he has criticised other theories, the theory he puts forward is more powerful or useful. If this theory can be destroyed, something will have been gained. It does seem, however, that perceptual processes, and the ways in which values are communicated, are useful concepts for research into problems of behaviour, including deviant behaviour in this country.

After the Crime

No matter what action may be taken to prevent crime, it is certain that some persons will continue to commit offences and that society will have to take some action to protect its members. A diagram (figure II) illustrates in broad outline the procedures which may follow the committal of a crime. In respect of one form of treatment—Borstal training for which research information is available—the results of treatment are shown. Roughly one in three of the offenders dealt with in ' open ' institutions committed further

Figure II


CRIME
DETECTED
UNDETECTED
POLICE ACTION
NO POLICE ACTION
PROCEEDINGS BEGUN
CAUTION
WRITTEN OFF
COURT ACTION
UNFIT TO PLEAD
NO CASE
GUILTY
NOT GUILTY
ABSOLUTE DISCHARGE
CONDITIONAL DISCHARGE
FINE
PROBATION
APPROVED SCHOOL
BORSTAL
PRISON
DETENTION CENTRE
OTHER
RECEPTION CENTRE (ALLOCATION)
TRAINING IN OPEN BORSTAL
TRAINING IN CLOSED BORSTAL
35%
60%
65%
40%
SUCCESS
ROUND AGAIN
ROUND AGAIN

crimes within three years, whereas approximately two out of every three of those dealt with in ' closed ' institutions were similarly re-convicted. Clearly the allocation procedure will tend to ensure that ' better material ' is sent to ' open ' conditions. The problem is, whether open treatment is better than closed for the purposes of securing reform, having regard to the type of offender normally trained in the different conditions.

Mannheim and Wilkins[15] were concerned with this problem and attempted a solution by means of ' prediction methods.' These methods enable measurements of the nature of the ' input ' to be made and compared with the results of the ' output.' This is done by setting up a ' discriminant theory model ' at the stage immediately following the sentence of the courts, that is, before the offender is received into the reception centre. At this stage a quantity of information is available which does not, of course, include any reference to the type of treatment to be received.

It is possible to express the ' pre-reception ' probabilities of recidivism as a linear function. This function provides a description of the ' input ' material, and enables that part of the difference in outcome which may be inferred to be due to the nature of the ' input ' to be partialled out. In this particular case, it was found that nearly all the variation in the outcome could be explained by the nature of the ' input ' as described by the equations. That is to say that the outcome of treatment was related almost entirely to the individual and his characteristics rather than to whether he went to a closed or open Borstal. There was, however, a small residual (which was just significant at the 1% level with a sample of 700) in favour of open treatment. This does not mean that open treatment was indicated for all cases—that all Borstals should be open institutions. It is possible to explain the difference in favour of open treatment in terms of an interaction factor which may be of various forms, and which, because of the design of the study, cannot be distinguished from factors of error inherent in the experiment itself.

The general power of the prediction method, however, may be appreciated from the fact that it was possible, using only information available before committal to Borstal, to forecast the outcome of treatment with remarkable accuracy. At one end of the scale, a group of offenders could be identified who were subsequently successful in 87% of the cases, whereas at the other end of the same scale a group with a success rate of only 13% was similarly identified. The main function of the method, however, is to provide a research tool. The efficiency of the tool depends upon the quantity and the quality of the information that can be obtained and used in its construction. Inadequacy of information is the main obstruction to developments in this area of research.

So far the results of studies of this kind, in this country, have failed to show that any one treatment is very much better or very much worse than any other. None of the results can be said to be conclusive. This applies to other methods applied to solving the same problem. A study by the present writer[16] of a matched sample of offenders placed on probation and other offenders dealt with in other ways showed no significant difference in outcome. The report of this enquiry concluded with the statement, " The negative result of this study is challenging. Why did undoubtedly different treatment policies make no difference in the subsequent criminal activity? "

There are several hypotheses which might explain these results. Some relate to the validity of the method of research used. But this seems a doubtful explanation. No solution using mathematical models can, of course, be as convincing as direct experimental methods. In this country experiments involving randomisation of treatment have not been made. In California[17], however, persons placed on parole were released to parole officers having case-loads of from 15 to 90 and some offenders were released three months ahead of their scheduled time. Parolees were allocated at random to parole officers with 15, 30, 60 and 90 case-loads, and whether an offender was released early or not was also randomly determined. A follow-up over two years failed to reveal any difference in outcome.

It is fair to summarise research work to date in the area of treatment as showing that there is no generally good treatment—a treatment which is suited for all types of cases. It seems that it is necessary to examine the interplay between offenders and treatment, rather than to consider treatment as a single variable. The ' treatment ' indicated for different types of offenders may be contra-indicated for others; the problem becomes who should do what, with what, and to whom!

If delinquency is seen to be an ' illness,' then it is necessary to find the nature of the different types of ailment that have so far been classified together. It is now clear that what is meet treatment for one offender is poisonous treatment for another! But as yet only one research has been able to suggest ways of differentiating offenders according to the type of treatment indicated or contra-indicated. This is a recent and costly project carried out by the Office of Naval Research[18] in the U.S.A. Offenders were most intensively studied and classified in many ways, but in particular according to measurable personality characteristics related to the ways in which they *perceived* the world—their reaction to problems of social interaction.

Some were classified as ' socially mature ' others as ' socially immature,' after thorough testing. A rigorous experimental design was used. The classified offenders were placed in three treatment groups, of which two (T1 and T2) made use of intensive ' living group ' therapy; the third (S) was the more usual training of a Naval Correctional Establishment. The results were tested by follow-up. The percentage successes in the three treatments revealed no differences between treatments except when broken down by the type of offender classified by the social maturity scale. The following table shows the percentages not re-convicted:

Offenders Personality Type	Treatment Type		
	T1	T2	S
Socially mature	70	72	61
Socially immature ..	41	55	60
All	59	65	61

The 'S' type treatment did not, as expected, interact with personality differences, whereas the intensive case work methods did interact; the socially immature were *more* likely to recidivate after this treatment, which was, none the less a better treatment for other offenders. It is interesting also to note that the offenders in the experiment were classified in several other ways by intelligence and personality typologies but it was only the classification based on the conceptual model of *perceptual processes* that proved significant. This experiment represents a big step forward. It shows that something does make a difference. But there are still many problems; this study applied to military delinquents; it relates to American culture patterns, and the typology, although discriminating between treatment (by treatment and personality interactions), may be criticised on a number of grounds. Moreover, it is not completely demonstrated that the differences found were due to 'treatments' rather than to other things which just happen to individuals undergoing different treatments. We still know very little about the treatment of offenders, but the nature of our ignorance is beginning to be revealed, and this is encouraging.

Potential Delinquents

It may appear that to consider crime from the point of view of the treatment of offenders is to lock the proverbial stable door. However, a large proportion of crime is attributable to persons who have committed crimes before. If recidivism could be prevented the total volume of crime would be reduced to perhaps between a half and two-thirds of the present figures. This would be well worth while. From the point of view of research strategy, it also seems reasonable to spend considerable research effort on problems of treatment. It is very difficult (and frequently raises ethical problems) to obtain information from the population at large, whereas those who have committed offences represent, sometimes literally, a 'captive population.'

Again the treatment of offenders costs society a considerable sum of money. As an interim objective it does not seem unreasonable to consider how money can be saved even if, as yet, we see little promise of being able to save souls. A scientific break-through in one area may lead to a break-through in another area. Once a person has been convicted and found guilty of an offence by due legal process, it is reasonable that society should attempt by various means to help and reform him, and to this end it is ethical for research workers to seek information. To what extent it is ethical to seek the necessary information from persons who have not violated society's norms is a point of dispute. Even if it were possible to obtain such information and to be able to identify a person before he had offended, what action would it be reasonable to take regarding him?

In certain countries, it has been considered to be reasonable to identify high risk children and to attempt to prevent their becoming delinquent. This it is claimed is justified on the grounds of conserving scarce resources. Another system is to identify a high risk *area*, and to attempt social action in respect of the whole of the juvenile population in the zone. The latter strategy seems to have been accepted without much difficulty, whereas the former has resulted in conflict between groups of social workers equally concerned with the welfare of society but taking different views regarding basic individual freedoms and, thus, the ethics of different forms of action.

The Gluecks[19] claim to be able to predict delinquency and to identify among young school children with reasonable accuracy those who are likely subsequently to commit offences. Their claims are somewhat exaggerated, but nevertheless it must be admitted that it would be possible to identify potential delinquents by the use of prediction methods. Such identification could take place at the age of say, 8 or 9 years. Any such prediction methods, however, have two types of error:

(*i*) some children will be predicted to be potential delinquents who will not in fact subsequently become delinquent;

(*ii*) some children who, in fact, do subsequently become delinquent will not be identified.

Whether the two types of error are equal or not in terms of their social consequences does not seem to have been considered. The New York City Youth Board is at present engaged in an experiment to validate the Gluecks method of pre-delinquent prediction and to develop other instruments. This action has led to bitter public controversy between Youth Board and the Citizens' Children's Committee[20]. Confounded with the ethical objections are, in this case, technical objections relating to the two kinds of error involved. Large scale community projects, where a whole area is subject to social action are also currently taking place in Los Angeles, Syracuse, and Flint, Michigan, and possibly in other places. No ethical controversy seems to have arisen with these studies. The main result of interest so far is that no matter what else the social scientists have not discovered in the United States, it must be recorded to their credit, that they have certainly persuaded the authorities that such research should be paid for.

Such projects are extremely costly. The Federal Government agency, the National Institute of Mental Health, itself in one year disposed of $101,000,000 on social research, and of this the limited field of criminology received the sum of $1,600,000 of which about three-quarters was classified as going into fundamental research. In addition to Federal Government spending, individual States invest large sums in applied social research. If it is suggested that much of this money is wasted, it is claimed that a certain amount of waste may be necessary in order to provide a climate for the expansion of social sciences as a *national necessity*. If there is no money, or if money is available only for the top research workers, research cannot expand; one way to break the vicious circle, it is argued in Washington, is to put in more money than can be reasonably abosrbed at the present as an investment in the future.

Conclusions

Although very little positive evidence can yet be put forward as to the causes of crime or the treatment of offenders, many beliefs have been shown to be unfounded. There is a beginning of some integration of the various theories, and theories are better constructed than in the past. The practical consequences of some quite powerful theories can be stated and the mechanics for testing them are becoming known. The future looks hopeful. The next few years are likely to prove very exciting for social research even if not as rewarding as may be hoped.

References

1 *Criminal Statistics, England and Wales.* London: H.M.S.O.
2 HUFF, D., 1954, *How to Lie with Statistics.* London: Gollancz.
3 LODGE, T.S., 1953, " Criminal Statistics." *J. Roy. Statist. Soc.*, A (General) CXVI, 283.
4 WILKINS, L. T., 1961, *Delinquent Generations.* London: H.M.S.O.
5 *Daily Express.* 12.1.61.
6 *Daily Mail* (editorial). 12.1.61.
7 *Observer* (comment). 15.1.61.
8 SUTHERLAND and CRESY, 1955, *Principles of Criminology.* Chicago: Lippincott.
9 GLASER, D., " Differential Association and Criminological Prediction." *Social Problems*, Vol. VIII, 6.
10 COHEN, A. K., 1955, *The Culture of the Gang.* Illinois: Glencoe Free Press.
11 KITSUSE, J. I. and DIETRICK, D. C., " Delinquent Boys: a Critique." *Amer. Soc. Rev.*, Vol. 24 (2), 208.
12 DURKHEIM, E., 1951, *Suicide.* Illinois: Glencoe Free Press.
13 SROLE, as quoted in MEIER, D. L. and BELL, W., " Anomie and Differential Access to the Achievement of life goals." *Amer. Soc. Rev.*, Vol. 24 (2), 198.
14 FIEDLER, F., *et al*, 1961. Technical Reports, Nos. 3, 8, 9. Centre for Research in Social Psychology, University of Illinois: Urbana.
15 MANNHEIM, H., and WILKINS, L. T., 1955, *Prediction Methods in Relation to Borstal Training.* London: H.M.S.O.
16 WILKINS, L. T., " A Small Comparative Study of the Results of Probation." *Brit. J. Delinq.*, Vol. VIII (3), 201.
17 Special Intensive Parole Organisation. Reports of the Department of Corrections, California.
18 GRANT, D., " The Treatment of non-conformists in the Navy." *Annals*, March, 1960.
19 GLUECK, S. and GLUECK, E., 1950, *Unravelling Juvenile Delinquency.* New York Commonwealth Fund.
20 Press Statements and circulars issued by the New York Youth Board and Citizens' Children's Committee of New York.

THE DESIGN, CONSTRUCTION AND USE OF STANDARDISED TESTS —Part 2: The Interpretation of Test Scores

by D. A. PIDGEON

IN Part I of this series (Pidgeon 1961) an attempt was made to describe briefly the development of a modern standardised test—how it was constructed and how the scores on it were transformed to a meaningful scale and were subject to errors of measurement. Full appreciation of this aspect of the standardised test is clearly important, but it is even more important that the scores derived from such a test should be interpreted in the correct manner and this implies a clear understanding of what the scores mean in relation to what the tests are supposed to be measuring.

Attainment Tests

So far as attainment tests are concerned there would appear to be little problem—arithmetic tests measure attainment in arithmetic, reading tests, attainment in reading, history tests, attainment in history and so on. Even here, however, lack of appropriate knowledge can lead to misunderstanding. Apart from the necessity of appreciating the merits of the scale used for reporting the test scores and of being aware of the tests' errors of measurement, the precise relationship between the content of the test used and the previous syllabus covered by the pupils taking it, must be known before the scores derived from the tests can be correctly interpreted. A high score on a ' mechanical reading ' test does not necessarily imply high reading comprehension, and scores on a mathematics or history test, the contents of which only overlap in part with what the pupils have been previously taught, may be quite misleading if they are compared with norms derived from pupils whose syllabuses included all aspects of the subjects covered by the tests. Apart from reservations of this kind, however, modern standardised attainment tests can be accepted as being more scientifically constructed and hence more reliable versions of the more usual type of examination paper familiar to all teachers.

It is clear that the real problem of understanding the true meaning of test scores arises from the use of tests purporting to measure ' intelligence ' and it is the intention of this present article to attempt to clarify some of the misunderstandings and misconceptions that have grown up around the efforts of psychologists to measure this quality.

Assumptions underlying ' Intelligence ' test scores

It is unnecessary to enter into the various definitions of ' intelligence ' that have been proposed (see Vernon 1960). As with electricity it is possible to measure it, and put the measurement to practical use without knowing exactly the nature of what is being measured. Indeed, many of the difficulties that have arisen in the use and interpretation of ' intelligence ' tests stem from the acceptance of a definition coupled with the belief that tests are capable of measuring the ability so defined. The word ' innate ' for example

may be included in a definition of intelligence but a possibly misleading assumption would be made in assuming that what tests are measuring is also innate. This is not to suggest that intelligence (in the everyday use of the word) does not denote something, which, in part at least, is innate, or that it is unnecessary or unprofitable to pursue enquiries into what intelligence really is. In the interpretation and use of scores from tests of ability the practical issue, however, is to define fairly precisely, not intelligence as such, but the meaning of the score obtained.

Nowadays everyone is so well acquainted with the fact that people differ in intelligence, however the term may be personally defined, that the notion that a child may fail at his school work because he is being faced with a task ' beyond his capacity ' is one that is readily acceptable. A teacher who has spent many frustrating months trying to teach an eight year old child to read or master the process of multiplication, is well prepared to believe that, although most children of this age can do these tasks, they are beyond the capabilities, at any rate for the time being, of this particular child. Clearly these frustrating months could have been avoided if the teacher had had some means of ascertaining at the outset that this child was limited in his capabilities compared with other children, and it appears that the means have been provided in the shape of an ' intelligence ' test.

Consider the case of two children, both aged 9 years, and both revealing a very low reading ability—a standardised score, say, of 75 on a test of reading comprehension. It would clearly be of the greatest value to the children's teacher, especially if faced with a class of 40 other pupils, to know whether any special efforts on his part to teach the two to read are likely to be at all rewarding. He applies an ' intelligence ' test—non-verbal, since the children cannot read—and one obtains a standardised score of 100, the other 70. Being cautious the teacher gives a second test which confirms these results. The teacher then reasons that the first child apparently possesses average intelligence and that whatever the cause of his poor reading it is certainly not lack of ability. Efforts on the part of the teacher, then, to find the cause of the difficulty and apply remedial techniques may well result in success. The second child is, however, manifestly dull and there is no need to look any further. Although poor, his present level of reading is all that can be expected and it is unnecessary to take any remedial action.

At this point we should note the basic assumption implicit in this line of reasoning. We may agree that children cannot perform tasks beyond their capabilities; and this may be translated as saying that capacity must necessarily limit output, or to use a much quoted analogy—it is impossible to pour more than a pint out of a pint pot. So far the words ' capabilities ' and ' capacity ' have been used, although ' intelligence ' could be easily substituted, accepting any one of the many definitions that has been formulated for this entity. However, when we move from ' intelligence ' to a score on a test of ' intelligence ' we make the further assumption that ' intelligence ' tests are yielding a direct measure of this entity—that is, a measure of the ' capacity ' of a child. It is, in fact, strictly true to say only that the ' intelligence ' test used by the teacher is giving a measure of the children's ability to answer the non-verbal perceptual type of items of the test, in the same way that the reading test was giving a measure of their ability to read sentences with comprehension. Both are dealing with present *function*. Viewed in this way, the non-verbal test is no more yielding an indication of *capacity* (innate or otherwise) than is the

reading test itself. Hence, although the teacher may be praised for taking remedial action for the wrong reason with the first child in the example, we must surely condemn the assumption that the second child cannot also benefit from his special help.

The Mathematics of Test Scores

The important point being made here is that ' intelligence ' tests do not yield measures of capacity—they do not and *cannot* indicate the true limits of performance in actual attainment. All teachers are familiar with examinations; many, perhaps, are not so familiar with the fact that examining necessarily implies measurement. For all measurements it is necessary to have two statistics—a zero point from which to start measuring and a unit with which to make the measurement. In the measurement of length, for example, the zero point is in fact absolute zero—no length, and the unit is arbitrarily defined as the yard, metre or some other measure according to the country of origin. In the field of mental measurement a difficulty is experienced immediately—there is no absolute zero point. If an examination or test were made in which some children scored zero, it would always be possible to make another easier one on which the same children now achieved a positive, though low, score. Exceptions to this rule are possible, for example, in the area of spoken vocabulary, but for practical purposes they can be ignored. To overcome this difficulty it is usual to adopt as the zero point from which to start measuring, the mean or average of the scores obtained on the test or examination, and the score of any particular child is then referred to as a deviation from this average, above or below. It will readily be seen that the actual score which denotes the average will depend upon the difficulty of the items in the test and the performance of the children on whose results the average was calculated. For any given test the average will vary according to the level of performance of the children taking it. In most physical measurements, such as height or weight, comparability between any two independent measurements can be ensured by the use of the same units and by making the measurements from the same reference point—zero. For the scores on any two mental tests to be strictly comparable, they must bear the same relationship to their respective averages (i.e. similarity of unit) and the averages must be derived from the same, or very similar, samples of children (i.e. the same reference point).

We may now consider an ' intelligence ' test, given to a representative sample of children and yielding an average score which, on the familiar scale of standardised scores, would be made to equal 100. Similarly, we may consider a test, or series of tests, of attainment, given to the same or a similar representative sample of children, and yielding an average score which, on the same scale, would also be made to equal 100. The differences between the averages, by definition as it were, is zero, and this is the same thing as saying that the average of the differences is zero. In other words, if, for each child the difference between his standardised scores on the ' intelligence ' and attainment tests was considered, the average of these differences for all children would be zero. This means that to counterbalance those children whose attainment test scores fall below their scores on the ' intelligence ' test, there will be an approximately equal number of children whose scores on the attainment tests *exceed* their scores on the ' intelligence' test, and this implies, if the ' intelligence ' test is regarded as measuring capacity, an approximately equal number of

'over-achievers' as 'under-achievers.' Now, the concept of the 'under-achiever' is only too easily appreciated, but what is the educational or psychological explanation of the 'over-achiever'? Is this really a case of more than a pint being poured from a pint pot or is the notion that 'intelligence' tests measure capacity wrong?

The relationship between the scores on two tests

Another way of presenting the argument, but in rather more concrete terms, is to consider the relationship between the scores on any two tests. It will be readily appreciated that, if the correlation between any two tests is known, then, given a child's score on one, it is possible to predict the most likely value of his score on the other, or at least, to define, with reasonable confidence, the range within which his score is likely to fall. In statistical terms this range is usually taken as plus or minus twice the standard error of the difference between the scores on the two tests. If the correlation is perfect, then the range is zero—the score on the second test can be predicted with certainty, it is the same as the score on the first test. As the correlation diminishes so the range increases until at the other extreme, with a zero correlation, it is in fact the full range of test scores—a score on the first test giving no indication whatever of the likely score on the second.

Table I below gives the most likely value on one test (Test B) and the lower and upper limits of the range defined by twice the standard error, for a set of scores obtained on another test (Test A), for each of four different values of the correlation coefficient between the two tests.

It will be noticed that for high and low values of score on Test A, the most likely score on Test B is not the same value but in each case one which is nearer to the total sample average of 100. The reason for this is the phenomenon known as *regression* and it arises from the fact that low scorers, for example, on the first test, if they change their score on the second test have a much greater opportunity to change in an upward direction than in a downward one. The child who is bottom on the first test cannot occupy a lower position on the second, although he can improve. The opposite, of course, occurs with high scorers on the first test—they can more readily move downwards than upwards. It will be observed that the regression effect increases as the correlation decreases.

The meaning of the figures in Table I can perhaps be best illustrated by considering a group of children all obtaining a standardised score of 70 on Test A. If it is known that Test A correlates 0·6 with Test B, then the majority of these particular children would obtain scores on Test B clustering around 82. Some, however, will obtain a score as high as 106, some as low as 58. Indeed, 5% of them will obtain scores outside this range, that is above 106 or below 58.

So far these tests have only been referred to as A and B. A further meaning can be given to the figures if Test A is now considered to be a test of, say, mechanical arithmetic, and Test B, a test of reading comprehension. The correlation between two such tests will vary somewhat according to the nature of the tests and possibly the age of the children taking them, but the value will not depart radically from 0·7. All teachers know that individual children vary in their relative performances on these two subjects and will not be unduly surprised to find that some children who were average for arithmetic (i.e.

Table I

Range of Predicted Score on one Standardised Test from Different Levels of Score on another Correlated Test

Test A—MEAN 100; S.D. 15 **Test B**—MEAN 100; S.D. 15

corr. = 0·90
S.E. = 6·60 2S.E. = 13·20

Score on Test A	Score on Test B		
	Lower Limit	Most Likely Value	Upper Limit
70	60	73	86
80	69	82	95
90	78	91	104
100	87	100	113
110	96	109	122
120	105	118	131
130	114	127	140

corr. = 0·80
S.E. = 9·00 2S.E. = 18·00

Score on Test A	Score on Test B		
	Lower Limit	Most Likely Value	Upper Limit
70	58	76	94
80	66	84	102
90	74	92	110
100	82	100	118
110	90	108	126
120	98	116	134
130	106	124	142

corr. = 0·70
S.E. = 10·65 2S.E. = 21·30

Score on Test A	Score on Test B		
	Lower Limit	Most Likely Value	Upper Limit
70	58	79	100
80	65	86	107
90	72	93	114
100	79	100	121
110	86	107	128
120	93	114	135
130	100	121	142

corr. = 0·60
S.E. = 12·00 2S.E. = 24·00

Score on Test A	Score on Test B		
	Lower Limit	Most Likely Value	Upper Limit
70	58	82	106
80	64	88	112
90	70	94	118
100	76	100	124
110	82	106	130
120	88	112	136
130	94	118	142

scoring 100) obtained scores as high as 121 or as low as 79 for reading. They may well be concerned to improve the reading of those scoring 79 or even to improve the *arithmetic* of those scoring 121.

A somewhat different interpretation is often employed, however, if, instead of arithmetic, Test A was called an ' intelligence ' test. The children scoring 79 on the reading test are clearly ' not working up to capacity,' while the scores of 121 would probably be viewed with suspicion and presumably no effort would be made to improve the *intelligence* of the children obtaining them. It must be remembered that the figures given in Table I are based on empirically determined evidence of the correlation between two tests. It is true that a verbal ' intelligence ' test may correlate higher with reading than 0·7, but the value is unlikely to be much above 0·8. What is important is that the figures apply whether both tests measure attainment or whether one is concerned with ' intelligence ' and the other with attainment. There is clearly no evidence that the scores on an ' intelligence ' test possess any special attribute to distinguish them from those on an attainment test, or to suggest that they can imply a limit to attainment test scores.

The concept of ' under ' and ' over ' achieving

Enough has been said by now to make it clear that the nature of mental measurement is such that it is impossible for a score on an ' intelligence ' test to indicate the limit of score that could be achieved on attainment tests. However, the concept of ' under-achieving ' is undoubtedly real and it is interesting to speculate why the belief that the scores from an ' intelligence ' test would reveal the under-achievers, should have been accepted in the face of evidence that the same scores would also reveal the anomalous over-achievers. The answer in all probability is that the latter were never found because they were never looked for.

Historically, the growth of the psychology of individual differences in the latter part of the 19th century and first part of the 20th and, particularly in this country, the stress placed by Spearman, Burt and others on the differences in inherited ability, lead naturally to such statements as Burt's (Burt 1937):

> " Intelligence, as psychologists use the word, is almost synonymous with what the Board of Education, somewhat uneuphoniously, has called ' educable capacity.' Capacity must obviously limit content. It is impossible for a pint jug to hold more than a pint of milk; and it is equally impossible for a child's educational attainments to rise higher than his educable capacity permits. In theory, therefore, we should always expect his educational age to fall below his mental age. In practice, however, when standardised tests are applied, the former is found to exceed the latter in a few rare instances."

If ' intelligence ' tests were measuring ' educable capacity ' then over-achievement was virtually impossible—and why look for the impossible? It is interesting, however, to speculate again why ' over-achievers ' did not turn up, even though they were not looked for. The answer is, of course, that a few were found, but were attributed to faulty testing or to some exceptional circumstance. The reason why larger numbers were not revealed was probably due to comparisons being made on tests standardised on different samples, and to the natural concern of educators for the dull and backward child.

The early ' intelligence ' tests described performance in terms of mental age—a concept that is easily understood by teachers and others and indeed is still used today. Now, chronological age is a measure on an absolute scale (that is, like length, it is measured from an absolute zero point—no age) and because mental age is described in the same units (years and months) it was easy to assume that it was also an absolute measure and hence that the difference between chronological and mental age or the ratio of mental age to chronological age, the I.Q., was also absolute—that is, had the same meaning in all circumstances. But mental age is no more than a raw test score and is subject to the same variations as any other test or examination scores. A mental age of 8 years is no more ' real ' or ' absolute ' than an examination score of 45%. However, the belief that it was led to comparisons being made between different ' intelligence ' and attainment test scores which in reality were not comparable. A mental age of 10 years, for example, was thought of (and unfortunately still is by many people) as being as ' real ' as a chronological age of 10 years, instead of being approximately equal to the average performance of the sample of 10 year old children on whose scores the test was standardised. Similarly, attainment ages were regarded as being ' absolute,' instead of being related in the same way to the sample on which the tests were standardised.

It follows, therefore, that only if the two standardisation samples employed were closely similar can comparisons in fact be made between the ' intelligence ' and attainment scores. If the standardisation sample used for the ' intelligence ' test included rather more below average children than did the attainment test standardisation sample, then the scores subsequently derived from the ' intelligence ' test would tend to be above those from the attainment test. And this situation often occurs. In the construction and standardisation of individual ' intelligence ' tests for example, deliberate attempts are made to produce norms from samples fully representative of the total population and this means the inclusion of the correct proportions of children who are educationally subnormal or even feeble minded (see e.g. Wechsler, D., 1949). But attainment tests are invariably standardised on representative samples of normal school children and the very dull tend to be automatically excluded.

It must be admitted that such differences in standardisation are unlikely to be very large and an examination of the differences between ' intelligence ' and attainment test scores for a large group of children would reveal a sizeable minority of ' over-achievers,' although certainly not equal proportions of ' over ' and ' under ' achievers. (See example in Table XIX, page 177, Burt 1921). However, such examinations are rarely carried out since attention is, for obvious reasons, more usually directed to backward children, or those thought to be in need of remedial education. Now children referred to remedial centres, say, for reading, form a selected group — at the very bottom of a distribution for reading. If these children are now given an ' intelligence ' test because of the effect of regression, their scores on this test will tend to be higher and most of them will appear to be ' under-achievers.' Indeed, if there are also errors of standardisation it would not be surprising to find that ' over-achievers ' are very rare. It will, of course, be appreciated, that if children at the upper end of the distribution for reading had been given an ' intelligence ' test, a large number of ' over-achievers ' would have been found.

Constancy of the 'I.Q.'

It is not the purpose of this present article to enquire into the various reasons why measured intelligence should have come to be thought of as a fixed entity that is largely inherited. An excellent account of the development of the 'capacity theory of intelligence' has been given by Daniels (1959). It is relevant to note, however, that the idea of the constancy of the I.Q. stems directly from the belief that 'intelligence' tests are measuring innate capacity. Regarded in this way, intelligence must remain constant and therefore variations in test scores or I.Q. can only be due to deficiencies in the test used. That 'intelligence' test scores do not remain constant can be clearly demonstrated by reference to the figures in Table I; the correlation between the same or very similar 'intelligence' tests, given with intervals of a few years, is usually no greater than ·8 and drops well below this value when the interval is increased or when different kinds of 'intelligence' tests are used (Vernon 1955A). Constancy of the I.Q. implies, of course, that correlations approaching 1·0 should be obtained with any interval of time. It is quite clear that 'intelligence' test scores are affected by 'environmental' factors to no small degree and numerous researches have been carried out demonstrating this. (See bibliography in Clarke & Clarke 1958).

What are 'intelligence' tests measuring?

It may seem to be a legitimate conclusion from the above account that if 'intelligence' tests are not of any value for measuring 'educable capacity' they have no use at all. To anyone accustomed to the 'capacity' theory of intelligence and who has attempted to use the difference between 'intelligence' and attainment test scores to reveal the 'under-achievers' such a conclusion may appear inevitable. Certainly it is the present writers opinion that it is better to abandon completely the use of such tests rather than to use them in the wrong way. Far better that attempts should be made to discover the difficulties of all non-readers and to explore every possibility of teaching them to read without recourse to 'intelligence' testing, than to abandon all efforts for some of them because they apparently reveal an innate dullness by obtaining very low scores on 'intelligence' tests.

But such resignation is not justified. The point was made earlier that it is unnecessary to define intelligence in order to use 'intelligence' tests; it is necessary, however, if the tests are to be used correctly, to understand a little more clearly what they are measuring. It is suggested in order to do this that an operational viewpoint be adopted or, in other words, that the score from an 'intelligence' test be interpreted in terms of what the test contains or in terms of empirically determined facts known about it.

This is not just a re-statement of the old definition—'intelligence is what intelligence tests measure.' In fact, it is desirable at this stage to sever completely the connections between what is understood by 'intelligence' and what is measured by 'intelligence' tests and it would probably help in this necessary separation of ideas if 'intelligence' tests were not so called.

The view taken here is that all tests measure intelligent behaviour whatever titles they may be given. A child answering the items in a reading 'attainment' test or in a

mechanical ' aptitude ' test or in a non-verbal ' intelligence ' test, is displaying intelligent behaviour albeit different aspects of it. And each of these aspects has its own importance.

Thus it should be clear that a child who obtains a standardised score of 70 on a non-verbal test involving the ability to discover relationships between diagramatically presented material, should not be classified as ' unintelligent ' even though it may be called an ' intelligence ' test. Rather, he should be noted as being, on the occasion tested, decidedly below average (in fact only better than the bottom $2\frac{1}{2}$ % of the sample on which the test was standardised) on a non-verbal test involving the ability to discover relationships between diagramatically presented material. Whether this knowledge is of any use to the teacher obtaining it depends upon what evidence there is relating the scores on this test to something meaningful to the teacher. It may be mentioned in passing that, so far as most ' non-verbal ' tests are concerned, there is very little evidence indeed to show that they have any value at all outside of research. (Vernon 1958).

But for other kinds of ' intelligence ' test—especially those involving items of ' verbal reasoning '—evidence of the kind required is available. There is, for example, a large body of research showing that the ' verbal ' or ' verbal reasoning ' test is the best *test* predictor, at the age of 11, of secondary school success. (Emmett 1942, Rutter 1950, Wrigley 1955, Yates & Pidgeon 1957). It is interesting to note the reasoning employed in describing the use of this test in 11+ selection. It is *not* said that because a child obtains a high score he is intelligent and will therefore succeed, or that a child with a low score is unintelligent and will therefore fail, but that there is empirical evidence to show that children with a high score stand a relatively high chance of succeeding, and children with a low score stand a relatively small chance of succeeding. No reference is made to the intelligence of the children and moreover this view enables a perfectly rational explanation to be made of low scorers who do eventually succeed.

Care should be taken not to regard this evidence of the high predictive value of verbal tests at the age of 11, as evidence of the value of other ' intelligence ' tests for predicting future attainments. Recent follow-up studies carried out at the Foundation clearly show, for example, that the best predictor at age 8 of reading at age 10, is a reading test (correlation = ·8). Non-verbal tests at 8 only correlate about ·5 with reading at 10, and moreover no significant relationship is found between non-verbal test scores and improvement in reading.

The demonstrated success of the 11+ verbal test probably arises from the fact that the ability required to answer it successfully has a great deal in common with the ability required for working successfully in most school subjects. In other words, the average correlation between verbal tests and all other attainment tests is higher than that between any one attainment test and all others. Thus, if it is required to put children into relatively homogeneous groups, e.g. for streaming, this can most easily and successfully be achieved by using a verbal test. There will, of course, be considerable overlap between ' streams ' formed in this way, but it will be less than if the streaming had been carried out be means of attainment tests alone. The overlap to be expected is not inconsiderable; Vernon (1955B) has estimated that in a three-stream primary school, roughly 10% of the children would require up-grading or down-grading a stream every year. (See also Daniels 1959).

The fact that verbal tests correlate relatively highly with general attainment also means that they are useful where it is required to match or compare different groups of children. This has specific value in research, but a teacher in a new school may wish to know how his present children compare with those he has been teaching. An examination of the mean and distribution of scores on a verbal test given to both groups will tell him this quickly, although more useful information could eventually be obtained by giving separate attainment tests.

In demonstrating the inadequacy of the ' educable capacity ' theory of intelligence testing, a certain stress has been laid on the similarities between ' intelligence ' and attainment tests, in particular that both kinds of test are measuring learned intelligent behaviour. It may, however, be argued that there are also certain basic differences, for example that while attainment tests are measuring the intelligent behaviour that has resulted from direct teaching, ' intelligence ' tests attempt to measure what has been learned indirectly. The difficulty here is that this is by no means a clear distinction; arithmetical problems are to be found in ' intelligence ' tests and typical ' intelligence ' test item types, such as analogies and classifications are often to be found in attainment tests (Pidgeon & Yates 1957). Moreover, what one child has not been taught directly, another may well have been.

Perhaps a better distinction, and one which is more in keeping with the operational viewpoint being suggested here is the difference in the area of intelligent behaviour covered by the two kinds of test. It can be supposed that there is an infinite variety of situations in which intelligent behaviour is demonstrated. (See Brimer 1956). Attainment tests cover only very specific areas or groups of ' situations ' (e.g. the items in a reading comprehension or mechanical arithmetic test), while for ' intelligence ' tests a wider area is covered. The area covered by group verbal tests is nevertheless somewhat limited and although it overlaps considerably with school attainments (giving rise to the relatively high correlations mentioned earlier), it overlaps much less with other areas of intelligent behaviour. Individual ' intelligence ' tests may be said to ' sample ' the ' infinite variety of situations ' to an even wider extent, but even so, the adequacy of the sampling is dependent upon cultural determinants. In other words, the content of any particular ' intelligence ' test constructed for a Western culture, will only contain a sample of the intelligent behaviour of that culture and will be useless for cross-cultural comparisons (Vernon 1955A). Even within a culture, there may exist differences of opinion on the adequacy of the ' sampling,' since, being infinite, it is impossible to define precisely the population from which it is drawn.

Viewed in this way, a score on an ' intelligence ' test could be regarded, but only with due caution, as indicating the average level of performance in the wider area of intelligent behaviour covered by the test; and being an average it cannot indicate the score to be expected in any specific (attainment) test in that area any more than a knowledge of a batsman's average in cricket predetermines his score in his next innings. It should be appreciated, moreover, that averages can be improved by improving component scores, whether the latter were originally high or low. While it is dangerous to pursue analogies like this too far, it can be observed that there is no law which says that component scores should ideally be equal to the average.

Theoretically there may be an upper limit to the possible score that could be achieved by any child on any test, but this limit is unknown (and, with present knowledge at least, unknowable) and since obtained scores must either be equal to this theoretical limit or below it, the only safe assumption is that it is possible to raise *all* scores. It must be realised that an ' intelligence ' test score no more estimates maximum potential than does an attainment test score; *both* are subject to the influence of factors likely systematically to depress them below a theoretical maximum. In other words, if, by effort of improved instruction, the level of intelligent behaviour called for in specific attainments can be raised, then it is an equally acceptable viewpoint that the area of intelligent behaviour covered by ' intelligence ' tests can also be raised. If a low attainment score relative to an ' intelligence ' test score is interpreted, therefore, as indicating the need to raise the attainment concerned, so also, must a low ' intelligence ' test score relative to an attainment score, be interpreted as indicating the need to raise the ' intelligence '—or more specifically perhaps, to improve the kind of intelligent behaviour covered by the ' intelligence ' test. And it is not meant by this that children should be coached to answer ' intelligence ' test items any more than they should be coached to answer attainment test items. Rather, they should be taught to act and think intelligently in all situations, in or out of school.

In conclusion it might not be inappropriate to sound a note of warning. The belief that scores on ' intelligence ' tests can indicate the level of performance to be expected in attainments may well have the effect of keeping the attainments of some children below what they could be, for ' intelligence ' test scores may be interpreted as ' expectancy ' scores and once a child is found to be ' working up to capacity ' there is a danger that no further efforts will be made to improve his attainments. Children at all levels of ability and attainment will be affected by this complacency, but the danger is perhaps greatest with the dull and backward, for dullness (i.e. low ' intelligence ' test scores) becomes interpreted as a *cause* of backwardness instead of being merely concomitant with it.

References

BRIMER, A., 1956. " The Relationship between Ability and Attainment—A suggested Revision of Current Theory." *National Foundation Bulletin*, No. 8.

BURT, C., 1921. *Mental and Scholastic Tests.*

BURT, C., 1937. *The Backward Child*, 1st edit., U.L.P.

CLARKE, A. M. and CLARKE, A. D. B., 1958 Ed. *Mental Deficiency: The Changing Outlook.* Methuen.

DANIELS, J. C., 1959. " Some effects of sex segregation and streaming on the intellectual and scholastic development of junior school children." Unpublished Thesis. Nottingham University.

EMMETT, W. G., 1942. *An Inquiry into the Prediction of Secondary School Success.* U.L.P.

PIDGEON, D. A. and YATES, A., 1957. " The Use of Tests in the Classroom—Aptitude, Intelligence and Attainment." *National Foundation Bulletin*, No. 9.

PIDGEON, D. A., 1961. " The Design, Construction and Use of Standardised Tests—Part I." *Educational Research*, Vol. III, No. 2.

RUTTER, D., 1950. " An Inquiry into the Predictive Value of Grammar School Entrance Examinations." *Durham Res. Rev.* I, 3-11.

VERNON, P. E., 1955A. " The Psychology of Intelligence and G." *Bulletin of British Psychological Society*, No. 26.

VERNON, P. E., 1955B. " The Assessment of Children—recent trends in mental measurement and statistical analysis." *Studies in Education*, No. 7. London Inst. of Educ. and Evans.

VERNON, P. E., 1958. " A new look at Intelligence Testing." *Educational Research*, Vol. I, No. 1.

VERNON, P. E., 1960. *Intelligence and Attainment Tests.* U.L.P.

WECHSLER, D., 1949. *Manual to Wechsler Intelligence Scale for Children.*

WRIGLEY, J., 1955. " The Relative Efficiency of Intelligence and Attainment Tests as Predictors of Success in Grammar Schools." *British Journal of Educ. Psych.*, 25, 107-116.

YATES, A. and PIDGEON, D. A., 1957. *Admission to Grammar Schools.* Newnes.

Research Note:

TEACHERS' VIEWS ON PRIMARY SCHOOL STREAMING

by Brian Jackson

Institute of Community Studies, London

THERE has recently been increased discussion of streaming in primary schools, and useful articles with good bibliographies have appeared. But we are still very frequently referred to Australian or American research because of the shortage of British studies, and of course this comparison of schools enclosed within quite different cultures becomes delicate, difficult, and sometimes impossible to make. In the present situation even fragments of fieldwork in British schools can be valuable, and this must excuse the incompleteness of the present note. It reports a very small piece of work done for the Institute of Community Studies, and is merely preparatory to a fuller enquiry into primary school streaming.

It seemed a useful thing, before we defined this final project, to seek a few more facts and ask some teachers' opinions on the subject. We therefore sent out 50 questionnaires to primary school teachers. The sheets were distributed for us by the National Foundation in Surrey, by the Nottingham Institute of Education, and the Leicester Department of Education. It seems reasonable to presume that amongst teachers in contact with these centres there was an unusually large proportion who were sceptical of streaming—and these results may be biased accordingly. No claims are made for the finality of the findings at any point: they are offered only as hints and suggestions—openings that might later be more strictly investigated.

Of the 50 class teachers, 11 taught A streams, and 10 taught B streams. 16 taught mixed ability groups in schools too small to stream, but large enough to provide one class for each age group. The remaining 13 took classes in schools so small that not only was streaming out of the question, but they had to teach children from different age groups together. There was no example of a school large enough to stream, but which did not, in fact, stream.

The teachers were first asked to place the children in one of five social groups according to their father's occupation. Doubtful occupations were described and allocated according to the Registrar-General's Classification of Occupation. The results are set out in Table I.

Table I

Father's Occupation	A Stream	B Stream	Mixed ability classes in small schools	Mixed age classes in very small schools	Social composition of surrounding local population
Professional and Managerial	21%	6%	12%	8%	6%
	37%	21%	25%	15%	24%
Clerical	16%	15%	13%	7%	18%
Skilled Manual	31%	24%	25%	44%	52%
	63%	79%	75%	85%	76%
Semi-Skilled Manual	21%	27%	28%	25%	16%
Unskilled Manual	11%	28%	22%	16%	8%
Total %	100%	100%	100%	100%	100%
Total Nos.	456	391	445	400	771,900

The most striking difference is between the A and B classes. If we call most manual workers 'working class,' and the rest of the population 'middle class,' then we have a rough guide that is generally understood. In this sense then there are 37% middle class children in the A streams, and only 21% in the B streams. Moreover, within the middle class those children with professional fathers tend even more to be in the A stream (21%) than the B stream (6%). This tendency goes all the way down: there are *more* skilled workers' children in the A class (31%) than in the B class (24%), and fewer labourers' children (11% as compared with 28%).

The small unstreamed schools provide some contrast to this. Again the number of middle class children is less, especially in the very smallest schools. It may be that the middle class is very ready to use the large streamed schools where its children generally get into the A class, but more reluctant to use the smaller, and most especially, the village schools. Either they don't live in these areas, or they send their children elsewhere. The final column shows the distribution of occupied males in the districts in which the questionnaire was used. This almost coincides with the general social position reflected by the unstreamed schools, and further illuminates the social selection within the larger streamed establishments.

We might, then, say that were this evidence to be borne out, streaming (whatever its educational reason or functioning) reflects some form of social selection within local society.

The teachers were next asked to place their pupils in three age groups. The results are shown in Table II.

Table II

Date of Birth	A Stream	B Stream	Mixed ability classes in small schools	Mixed age classes in very small schools
September 1st to December 31st	38%	28%	33%	32%
January 1st to April 30th	36%	37%	34%	34%
May 1st to August 30th	26%	35%	33%	34%
Total	100%	100%	100%	100%

In the 29 classes from smaller, unstreamed schools the distribution by birthday is fairly even. But both the A and B classes break this pattern in opposing ways. In the A classes 38% of the children were born between September 1st and December 31st; in the B classes only 28%. There was little difference for children born between January 1st and April 30th, but the scales dipped against those born between May 1st and August 30th. In other words the oldest children in the year were tending to go into the A classes and the youngest into the B classes. Again, we might say that, were this evidence born out, then (whatever the academic reasons, judgements, or measures used) there was some tendency for streaming to become a mild form of *selection by age within a year group.*

It is interesting to speculate whether this is not directly connected with the times at which the children entered infant school. Since children usually enrol in the term in which they become 5 years old, this means that some children may spend six terms in the infant school whereas others spend only 4. Or to put it more dramatically, one child may have had 50% more schooling than another when, at the age of 7, he is assessed for an A, B or C class.

If so much does depend on the *original* judgement of a child's capabilities as A, B or C, then it would be useful to know how the streamed schools first assessed their children. Of the 21 streamed classes, 2 had already been streamed from the infant stage, 8 had been streamed at the age of seven, and 1 had waited until the age of eight. Table III sets out the methods used.

Table III

Streamed at 6		
(*i*) Arrived streamed from infant school	2	Main weight falls on home background plus successful start in the infant school.
Streamed at 7		
(*i*) Streamed according to infant school 'report'	6	
(*ii*) Streamed according to infant school 'report' plus an 'examination'	2	
(*iii*) Streamed according to school attainment test in English, Arithmetic and Reading	6	
(*iv*) Streamed according to school attainment tests in English, Arithmetic and Reading, plus a declared intelligence test ..	1	Some weight given to potential as opposed to attainment.
(*v*) Streamed according to declared external attainment tests in English, Arithmetic and Reading, plus a declared intelligence test	2	
(*vi*) Streamed by using a declared intelligence test only	1	
Streamed at 8		
(*i*) Streamed according to first year teachers' recommendations ..	1	
Total	21	

Of the 21 classes, 10 placed a heavy reliance on the child's success and teacher's judgement in the infant school. 6 others streamed according to what the child had attained by the end of infant school and beginning of junior school, using methods and tests of their own. Only 4 schools made a measure of potential as opposed to achievement. Only 2 schools used a full set of recognised external tests on which to base their judgement. The overall tendency was for the school to stream children according to what home plus 4, 5 or 6 terms of infant school had given them, judged according to methods which were particular to that school rather than external and standardised.

If streaming in these 21 schools was somewhat affected by age and social background, if it took place so early and was determined so much by infant school success, then how did the teachers feel about it? Did they endorse it, or not? Table IV sets out the replies.

Table IV

Attitude to streaming	A Stream	B Stream	Mixed ability classes in small schools	Mixed age classes in very small schools	Totals
For Streaming	7	6	7	4	24
Qualified approval of Streaming	2	2	3	3	10
Against Streaming	2	2	6	6	16
					Total: 50

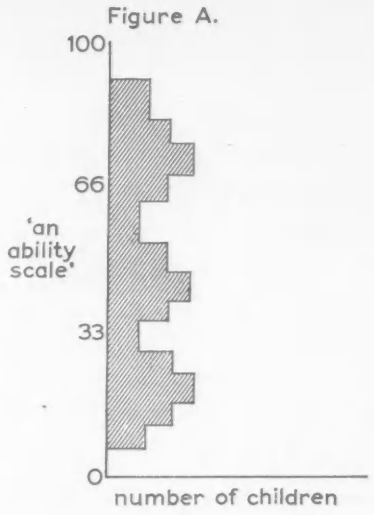
The largest group (24) were unequivocally in favour of streaming. But the number against was perhaps abnormally large (16), and there were 10 who defined some intermediate position (streaming only under certain conditions). One difference is between the teachers of streamed classes, and the rest. The majority of the former endorsed streaming (13 to 4), whereas in the small schools only a minority favoured streaming outright. It would be valuable to know whether this merely means that teachers taught in schools in whose organisation they believed, or whether it had to do with the direct experience of handling streamed or unstreamed children. For example, did the teachers in small schools draw their confidence in non-streaming directly out of their daily work with such classes? Had the streamed teachers ever tried (or seen) unstreamed teaching—or *vice versa*?

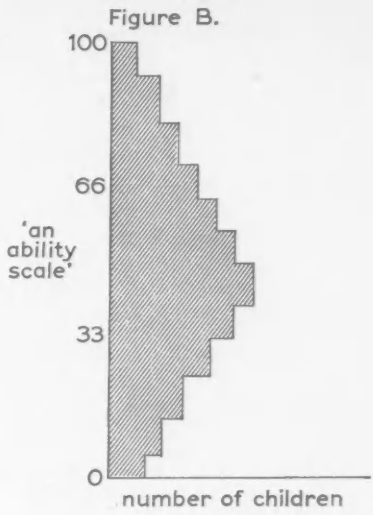
The 10 teachers who accepted streaming with some qualification are worth further examination for their indeterminate position revealed some of the many things that teachers weighed up as they made their decision, for or against. Many added the proviso " providing it is fluid streaming," but it is difficult to know how much more than a merely formal thought this was. Daniels has shown convincingly enough how a sample of teachers may imagine that there is very much more fluidity in a streaming system than in actual fact there is. The connections with the ' self-fulfilling prophecy ' that can be built into a streamed school are all too obvious.

Several accepted streaming until, in their opinion, the standards of primary school teaching were raised. If the quality of teaching improved then " good teacher—no streaming." They felt that we had not, at the present moment, enough teachers of sufficient calibre to manage unstreamed classes. Yet others felt that streaming was the only way to handle large classes, but that it might be abandoned when there was an improvement in the pupil-teacher ratio: " I endorse streaming under the present conditions of large classes. If classes were ever reduced to 30 then I would re-consider my decision." Others still felt it was only necessary in the years immediately preceding the 11 plus, and not at 7 or 6; and a few wondered if ' setting ' might help, but could find no way around ' time-table difficulties.'

It would seem then that different teachers may see streaming as serving very different ends; and it looks as if the streaming system is supported, in their view, by many interconnected pressures—the selection examination, the size of primary school classes, parental pressures, class rather than individual teaching, the quality of existing teachers, and even their hopes of promotion.

They were next asked, " *What exactly is the purpose of streaming?* " About half the teachers felt that there was a single *raison d'être* to streaming, but the rest suggested two or even three purposes. 22 teachers said quite simply that it was a device whereby children of similar ability were taught together: when they considered children in any one year they wrote as if they conceived of them as 2 or 3 separate ability groups, each clustering around its own mean, as in figure A. Two teachers disputed this picture, and added that the purpose of streaming was initially to " close the spread " at top and bottom so as to provide more homogeneous groups: they acted on the conceived spread of children's ability shown in figure B.





This is extremely tentative, but we might suggest that whereas the spread of children's ability is more likely to resemble figure B, the effect of streaming may be to change it to figure A. Many teachers however (especially in streamed schools) considered figure A to be the original fact and not merely the result.

16 teachers said that the function of streaming was to ensure that bright children were not held back, and so missed their 11 plus places. 8 of these teachers approved of streaming, and 8 objected to it. Those who saw streaming or non-streaming as a black and white choice agreed that it was paramountly a pre-selection device.

11 teachers, on the other hand, thought that streaming was not designed for the gifted child but rather to protect and help the backward. No A stream teacher thought this, and 9 out of these 11 replies came from schools in which no streaming took place. So teachers at a distance from streamed schools imagined it an egalitarian device, whereas those teaching streamed classes saw it as especially concerned with the preparation of grammar school children.

8 teachers said that streaming was essentially a device to make the teaching situation tolerable in large schools, and its first purpose was to make life easier for the teacher. These were mostly men and women who took up an intermediate position on the opposition of streaming and non-streaming. A further 6 teachers (of whom only one favoured streaming outright) said that streaming was a strategy to make the *head teacher's* life more tolerable—simplifying organisation, improving ' efficiency.'

There were, then, curious differences of opinion as to what streaming was for. Teachers engaged in handling streamed classes had fairly clear views, usually connected directly or implicitly with the 11 plus selection (" for the bright child to have the best

teachers "). But the majority of teachers at a distance from streamed schools discussed it as one of the settled ' Institutions of England,' and accounted for its continuing existence (as they might with such other Institutions as the House of Lords or the Royal Family) in a variety of often opposing ways.

The teachers were next asked, " What kind of people or groups in society support primary school streaming? " 16 gave a directly social answer. 11 of these said that it was supported by professional class parents using state primary schools. 3 said support came from the " aspiring middle class," and one from " the lower middle class—particularly when there are poorer children in the neighbourhood at the same school." 14 answered in slightly different terms, saying " A stream parents " and sometimes transposing this with " intelligent parents "—which implied perhaps that streaming not only selected children, but selected parents and families.

18 called on professional experience—10 of these answering that whoever objected to streaming, " the majority of teachers " were for it. 6 others simply added that endorsement came from those " who had had experience," and the remaining 2 narrowed it down to " good A stream teachers."

13 claimed that streaming was supported by those who cared especially for grammar school education, and 2 of them singled out head teachers as caring very much in this way. These comments did not usually come from supporters of streaming.

Other answers given were " traditional " or " *laissez faire* " elements in society, teachers only interested in traditional classroom methods, the " uneducated in education," and " most of western civilisation which is based on discrimination and competition."

So whatever teachers took to be the purpose of streaming most of them felt that, as a fact, it often worked to select and support *existing* groups within society, and to support teachers whose status, future, and self-respect had some connections with successful eleven-plus results.

The teachers were next asked, " What kinds of people or groups in society are sceptical of primary school streaming? " 13 teachers answered immediately in terms of parents with children in low streams, but thereafter the answers broke up interestingly, and very much in contrast with the previous question.

12 teachers said streaming was doubted by those with left wing political views. 2 teachers refused to answer because, they said, the question " was a political one." This compares markedly with the previous block—to doubt streaming was felt to be a " political " activity, but to support it wasn't.

10 teachers felt the doubts came from people working in education who were not actually teaching. 7 said " educational psychologists," and the others seized on " educational philosophers," " educational reformers," " sociologists " and " educationists." These remarks were invariably hostile.

There was no other clear block of answers, except that 6 A stream teachers (and no one else) said that opposition only came from " inexperienced " teachers. Otherwise the replies scattered widely and many of these answers were in the form of sharp criticism.

The criticisms were extremely striking. All the previous questions had been answered very coolly, but this one drew sparks—especially from those who endorsed streaming. Prose heated up and it was stated that opposition to streaming came from:

" those who *like to consider themselves* ' modern educationists' "

" those with a chip on their shoulders "

" free thinkers and nebulous characters "

" people who wrongly believe that men are born equal "

" socialists with inferiority complexes and/or dull children"

It was as if these teachers felt themselves to be on the defensive, holding a difficult position and able to deal with powerful opponents only after they had reduced them to caricature.

The teachers were finally asked, " What would happen if streaming in schools was abandoned? " 17 felt that there would be a " social gain "—this group included most of those who were opposed to streaming and, intriguingly enough, the majority of the A stream teachers. Only 1 teacher felt there would be a social loss.

Against this 13 teachers felt that general academic standards would be lowered, and a further 8 mentioned that it would be the " bright children " who would lose. 4 believed that the backward child would also lose. On the other side 8 teachers thought that the general standard would improve, and a further 5 that backward children especially would gain. No one claimed that non-streaming would improve the attainment of the most gifted children. All the A stream teachers, save one, confined themselves here as elsewhere to possible effects on A stream children, in a ' school within a school ' spirit. Other teachers discussed all children fairly equally.

4 replies suggested that life would now be harder for class teachers, but a further 5 claimed that the morale of B and C teachers would leap, for teachers also would be ' unstreamed.' Other responses included a marked increase in the demand for smaller classes, a new impetus to individual rather than class teaching (" the teacher would replace the instructor "), and a decrease in 11 plus pressure from parents.

The general position was that the largest group felt that unstreaming would result in a distinct " social gain." But this was not seen as the pre-requisite condition for the multiplication of ability, for most also felt that academic standards would be lowered. Teachers disagreed over which was the more important of these two emphases, and as one A stream teacher put it, non-streaming might give " a far more balanced society with perhaps some lowering of educational standards," *and so he was utterly against it.* Comments were always directed at the reaching of a certain academic standard at a certain age. The idea that children might eventually, under these conditions, reach or surpass a declared academic standard—but only after several more years at school or college—was treated as " lowering standards." That is to say, most teachers had built into their concept of ' standards ' an essential idea of *speed.* Fast growth was their ideal.

General

Some teachers added some general remarks of their own. Several wished to press the point that streaming was not just an educational device to be discussed in enclosed educational terms; it was an understandable impact on the schools of something basic to Western ideology. One pointed back to Aristotle: " Streaming takes place in all walks of life and the community. Read Aristotle! "; another to new Africa: " The emergence of new nations in Africa has demonstrated clearly the need for a minority of highly-trained professionals in any given country. To produce these I think some kind of separate and accelerated programme must be available in the primary school. *It is not clear that the remainder need streaming.*"

Altogether those who favoured streaming were much more consistent and detailed in their opinions. Their sharper organisation of ideas suggested that as a body they would be much more powerful in schools than the rest. They came from every kind of class and school, and held in common the belief that the function of the primary school teacher was to select and rapidly advance a gifted minority of pupils—for, even allowing for some social loss in the younger years, this was ultimately for the common good. Streaming for them was not just one schoolroom way amongst others, but in the end a response to basic values and drives in our society. They felt their position was under challenge, and at times maintained the temper of their argument only by protective caricature; but, for all that, they firmly held their ground with statements whose quality was that of the defiant trumpet-call.

This very small enquiry points to the possible effects of social class and age on primary school streaming, and reminds us of the layers of belief which underpin its practice. Yet it seems that teachers in schools where these effects are marked, are not distinctly aware of them, are unclear about what—besides ' children '—is actually being streamed. It would be an excellent thing if teachers in schools would more often analyse the children in the very simple ways of Tables I and II, and perhaps those who can, ought to encourage such activity. Whatever direction opinion or action takes, it cannot but be a good thing for teachers to know the facts about the classes they teach, and to be fully alive to the processes they encourage and serve.

References

DANIELS, J. C., " The Effects of Streaming in the Primary School." *British Journal of Educational Psychology*, Vol. XXI, Part I, February, 1961. (With a bibliography).

PAPE, G. V., " Accident of Birth." *Education*, November 16th, 1956.

RUDD, W. G. A., " The Effects of Streaming: A Further Contribution." *Educational Research*, Vol. II, No. 3, June, 1960. (With a bibliography).

YATES, A. and PIDGEON, D. A., " The Effects of Streaming." *Educational Research*, Vol. II, No. 1, November, 1959. (With a bibliography).

Research Note:

OVER 135 I.Q.

by JOHN NISBET and ALASTAIR GAMMIE

Department of Education, University of Aberdeen

WHERE Local Education Authorities have kept detailed records of the results of tests applied in selection for secondary education, there is a rich field of material for useful, short-term research studies. For example, Terman's classic study of gifted children in California could now be repeated, at least in its initial stages, in many areas of this country without undue expense or difficulty. Such follow-up studies would help to throw light on the question of how effectively our school system is providing for the highly gifted pupils and on some of the controversies associated with the proposed expansion of higher education.

The Aberdeen Survey

The records of the Transfer Examination (for transfer from primary to secondary school) in Aberdeen over the period 1950-53 were used to demonstrate the practicability of such a follow-up. This examination includes two group tests of intelligence or verbal reasoning each year. The scores in these tests are standardised within each age-group to give a mean of 100 and a standard deviation of 15 for Aberdeen pupils. We selected for our inquiry those pupils who scored an average of 130 or over on the revised quotients of these two intelligence tests. As the average score of Aberdeen pupils by national norms is usually about 104 or 105, the minimum level taken here is equivalent approximately to 135 I.Q. In the four years a total of 159 pupils obtained scores at this level or above. The total eleven-year-old age-group in each of these years was about 2,700.

It is possible that errors of measurement resulted in the inclusion in our group of some whose talents were not as outstanding as their scores suggested. However our use of an average score from two tests reduced the likelihood of such error.

The matriculation records at Aberdeen University for the period 1956-61 showed that 78 of these pupils had entered university. Only a very small proportion of students from Aberdeen schools go to other universities (except for post-graduate study). Thus, we can state with some confidence that only about fifty per cent. of these very able persons have gone to university—and this in a country and a city where the tradition of a university education for all who are able to qualify for it, is well established. In Terman's Californian group of approximately thirty years earlier, the proportion who went to college was nearer ninety per cent. What has happened to the fifty per cent. in Aberdeen who did not go on to university?

The 1953 Group

To answer this question, a single year-group was followed up individually. In the 1953 Transfer Examination, 36 pupils scored an average of 130 or above in the two intelligence tests. Of these, 20 are now at Aberdeen University. Two of the remainder are still at school (one having been retarded by a serious illness), but neither seems likely to reach university entrance standard. Two others left school without sitting the Scottish Leaving Certificate, one to become a nurse and one to enter the Services. Eight sat the Leaving Certificate but failed to get university entrance; two of these are working in insurance, two in journalism, and the other four are continuing further education, in technical college, art college, commercial college and teacher training college. Four others did sufficiently well in the Leaving Certificate to enter university if they wished, but decided otherwise: one holds a commission in the Services; one is in the Executive Branch of the Civil Service; one is an apprentice chartered accountant; and one a nurse.

This review of the occupations entered by those not at university shows that it would be quite wrong to classify this group as ' wastage.' The follow-up would have to be extended over many years before we could judge whether the initial promise had been fulfilled. The one who left school without a Leaving Certificate to join the Services, for example, was recommended for a commission within three months of enlistment. Others may yet distinguish themselves in other ways. A university education is not necessarily the best choice for all persons in this range of ability.

Fathers' Occupations

However, if we look at the fathers' occupations in this group, we find a trend which resembles that reported in the Ministry of Education (1954) study of *Early Leaving*. The proportion attending university is much greater among those whose fathers are engaged in professional and non-manual occupations than among those from homes where the father is in a manual trade or is an unskilled labourer (Table I).

Table I. 1953 Group: Fathers' Occupations

	Father's occupation: Non-manual	Manual	Not known	Total
At university	15	3	2	20
Not at university	4	10	2	16

The disproportion in the 1953 group would seem to be particularly marked. For the other years, only the fathers' occupations among those attending university were traced: Table II shows the distributions in these years.

Table II. 1950-53 Groups: Fathers' Occupations

Year	Those at university: fathers' occupations—				Not at university
	Non-manual	Manual	Not known	Total	
1950	12	5	3	20	22
1951	9	7	2	18	22
1952	11	7	2	20	21
1953	15	3	2	20	16

Conclusion

The purpose of this report is to demonstrate that a survey of gifted pupils can now be made from readily available records. The present survey showed that only half of the pupils who scored 135 I.Q. or over at age eleven proceeded to university. The area in which this survey was carried out was particularly well suited to such a follow-up, since virtually all the university students could be traced from the records of a single university. But much of the information was got directly by post or by phone, or through the secondary schools. In other areas, the records of local authority awards could also be consulted to trace the educational careers of pupils. It proved easy to gain the co-operation of the subjects in an inquiry on this topic: our description of the investigation as a follow-up of pupils whose primary school careers had been outstanding, seemed to dispel suspicion at once and established a favourable attitude in all those we approached.

Research Note:

TEACHERS AND THE SOCIO-ECONOMIC FACTOR

The Socio-Economic Factor as an Unstable Factor Liable to the Influence of Social Change

by E. J. GOODACRE

A recent study by Morris for the National Foundation on the teaching of reading found that of a number of variables the one most significantly related to Reading Attainment was socio-economic. This is not a novel finding, since other studies have stressed the importance of this factor. But what has always been interesting in such studies is the number of different methods that have been used in the last few decades to try to assess this elusive variable. Their variety tends to emphasise the complexity of the variable whose very name implies the wedding of two perhaps disparate aspects. Each of the two is likely to reflect social change, and the changing nature of the measures used may itself be due to a genuine change in what is to be measured.

Fleming, writing on this subject in 1943, pointed out that Socio-Economic Status has been assessed in a variety of ways: by relying on the opinion of representative townsmen or of the headmaster and attendance officer of the school; by the use of questionnaires; by gathering information as to taxes, possession of a telephone, the rent or size of the house, or the occupation of the family; by weighting reports based on material possessions and the appearance of the home; and by taking the school itself as a measure of social grading—a procedure justified by the observed fact that considerable homogeneity of social type is a characteristic of most schools.

In her own study, Fleming assessed the economic aspect and defined it in terms of material prosperity. She did however suggest that further investigations should consider evidence as to the more intrinsic characteristics of the homes and compare these with the results from measures more strictly socio-economic in nature. A little later, Burt (1947) commented that the Social Surveys in this country " have hitherto been limited chiefly to economic conditions and the material standards of life. The social attitudes, ideals and behaviour of different groups are barely touched upon."

Inquiries such as that into early leaving showed there were still differences between pupils at the secondary school level related to home conditions, and that it was possible that the provisions of the Welfare State, whilst obviating some of the strictly economic discrepancies between social groups, only revealed more subtle social differences. Floud, Halsey and Martin pointed out that whereas in the past, the problem of social waste in education could be seen in comparatively simple terms, the influence on the distribution of educational opportunity of the material environment at home and at school before the age of selection is tending to diminish. Because of social reforms and general prosperity, social factors influencing educational selection reveal themselves in more subtle forms today.

The most recently published contribution to this problem is that by Fraser. She studied school progress using an assessment of environmental conditions in economic, motivational and emotional terms. Her findings show clearly that such factors are more closely correlated with school progress than with intelligence, and that the difference between these correlations is highly significant. One of these environmental factors was that of parental encouragement which correlated +·66 with school progress. It is interesting that the rating of parental encouragement used was that made by the teachers; and Fraser makes the point that little is known about the basis on which such assessments are made. She suggests that in many cases it is unlikely that the teacher knows the parents and is consequently unable to assess their attitudes directly, and that in the majority of cases the teacher must be forced to make an assessment from indirect clues such as the help the child receives with homework from his parents.

Teachers' Consciousness of the Socio-Economic Factor

Undoubtedly teachers are likely to be aware of some at least of the socio-economic factors, but there is indeed little evidence available as to how and on what ' indirect clues ' their awareness is based or the effect of it on their teaching. The terms ' good ' and ' poor ' home background are commonly used in staff rooms, and sometimes they imply differences of expectation as to individual children's attainment and potential abilities— " Poor little thing, you can't expect much, she comes from such a poor home," or " you know, he should be able to do better, there is a good home there."

It is not clear in what sense such remarks are being made: whether they are purely descriptive of the material standards; or are meant to imply less concrete evaluations such as the degree of motivation and the general emotional climate of the home. We do not in fact know whether the reference is to the ' economic ' or the ' social ' or to some blanket two-dimensional factor. It is interesting that the evaluative term ' good ' is used rather than ' rich ', ' well off ' or ' prosperous ' which are applicable only to the material concept. ' Poor ' on the other hand can be used either in the strictly economic sense of ' having little money or means,' or of ' deserving of pity, unfortunate ' in the qualitative sense.

Recently in the course of a reading research project being carried out by the National Foundation in a large educational authority, the opportunity arose to find out more about what the terms ' good ' and ' poor ' mean to teachers when applied to their pupils' background. By means of a questionnaire, 206 reception class and head teachers were asked to say what in their opinions were the characteristics of a ' good '/' poor ' home background. Additional questions were asked to try to elucidate the basis of teachers' assessments of environmental conditions, that is to find out more about those ' indirect clues ' which must be the basis of such judgements and which Fraser suggested as worthy of future research.

There was little difference between the mean number of items mentioned for either category, an average of 2·23 characteristics for ' good,' and 2·95 (slightly more) for ' poor.' Perhaps it is easier to say what is wrong with a thing, than what is right with it. Head teachers tended to give slightly more items than class teachers, which may reflect their more frequent contacts with parents and rather longer experience.

Table I

'Good'/'Poor' Characteristics

(206 teachers in 100 primary schools)

Rank Order		Characteristics	Description	Good		Poor	
Good	Poor			Total No. Chars.	% of Total No. given	Total No. Chars.	% of Total No. given
1	2	Motivational	Parents interested/not interested in child's activities, spend time with child, answer his questions, mental provision/lack of, etc.	151	26·87	125	20·53
2	1	Emotional Climate	Stable/unstable home life, child loved/not wanted, broken home, family unity/lack of, mutual family consideration/bickering	131	23·31	128	21·02
3	5	Cultural Factors	Good conversation/swearing, poor speech, visits to interesting places/too much street play, right amusements/pools, dogs, etc., books/lack of	97	17·26	66	10·84
4	6	Material Provision	Good/poor food, clothing, plenty of/insufficient sleep, physical care/lack of, neglect	44	7·83	56	9·20
5	3	Moral Training	Firm but kindly discipline/laxity, little discipline, 'children run the streets,' spiritual, religious faith/lack of, independent/indulged	37	6·58	78	12·81
6	8	Regularity, organisation, training	Regular/irregular meals, sleep, habits; cleanliness/dirtiness; planned/unplanned television watching ..	32	5·94	35	5·75
7	4	Economic Factors	Housing/overcrowding, mother at home or makes provision/both parents working, financial security/insecurity, hire purchase	31	5·52	73	11·99
8	7	Interest in Child's School Progress	Parental help with school work/lack of, interested/not interested in education	30	5·34	36	5·91
9	9	Unclassifiable	Quiet/noise, tidy/untidy, over-anxiety, settled schooling	9	1·60	12	1·97
				562	100	609	100

As can be seen from the first of the two tables, the most often mentioned characteristics in the ' good ' category were motivational and emotional. Although there were a number of church schools represented in the sample, the concept was rarely defined in evaluative or spiritual terms. Of course, this may well be one of those instances of the respondent so taking an aspect for granted, that he or she omitted to mention it when confronted by an open ended question. When it came to describing the ' poor ' home, however, moral training and spiritual values were more to the fore, mainly in the sense that the ' poor ' home was that in which material standards were dominant to the detriment of spiritual development.

Difference According to Social Area

It was possible to divide the schools into three groups based on the differences of social class distribution for the areas in which the schools were situated. This information was obtained from the *Report on Greater London and Five Other Conurbations* (1956) based on the 1951 Census figures for England and Wales. Area A is that in which there is a greater proportion of professional occupations, whilst C is that in which the semi-skilled and unskilled occupations predominate. The second table shows the emphasis placed on the different aspects according to the social background of the school in which the respondent taught.

Table II

Percentage by category of Different ' Good '/ 'Poor ' Characteristics in each Social Area

(206 teachers in 100 primary schools).

Characteristics	Good			Poor		
	Area A	Area B	Area C	Area A	Area B	Area C
Motivational	25·99	25·77	28·38	21·16	19·64	20·63
Emotional Climate	25·99	23·93	20·72	24·34	23·21	17·06
Cultural Factors	14·7	20·25	17·12	8·47	10·71	12·70
Material Provision	8·47	6·75	8·11	5·29	8·93	12·30
Moral Training	7·34	6·75	5·86	16·40	10·12	11·90
Regularity, organisation, training	8·47	6·75	2·70	7·94	4·17	5·16
Economic Factors	2·82	5·52	7·66	9·52	13·69	12·70
Interested in Child's School Progress	5·08	4·29	6·31	5·29	7·14	5·56
Unclassifiable	1·13	0·00	3·15	1·59	2·38	1·99
	100	100	100	100	100	100

The mean numbers of items for each were similar, but some items were more often mentioned from some areas than others. For instance, in describing the ' good ' home background, the ' higher ' the social area, the more often items concerned with regularity were mentioned, or as might be expected, the ' lower ' the social area, the more frequent were the economic items. Some of the differences may represent differences in the various teachers' familiarity with or knowledge of the area in which they taught. For example, the more frequent occurence of the emotional aspect in the ' higher ' social area could be interpreted as a greater awareness by the teachers of the problem of family

unity in the particular area. In the ' lower ' areas, where divorce is perhaps less common than separation and alterations of family groupings, the ' broken ' home may make less impact on the child, and hence on the teachers' awareness.

What is evident from this breakdown by social areas, is that teachers in different areas are probably using the idea of ' good ' and ' poor ' home with different criteria in mind. The teacher in the C area may be thinking of background in economic terms whereas the teacher in the A area may be concerned less with those obvious material differences, and more with those ' subtle forms ' referred to by Floud, Halsey and Martin.

More characteristics of a motivational nature were reported in the C area, in regard to the ' good ' home background, and one is tempted to wonder whether the teachers in these areas are perhaps expecting co-operation with the school and an interest and ability to further school aims and methods, which may be difficult or impossible to achieve in such areas. For instance, in the specific matter of teaching reading, parents in such areas, experiencing overcrowding and a lower standard of living may be able neither to buy the books of the school reading series nor to help with ' hearing ' reading, which teachers reported as being the evidence of a ' good ' background.

Teachers' Attitudes to Home Background

Very little seems to be known about either the extent or the effect of teachers attitudes to their pupils' home background. It may well be that teachers differ greatly in their attitude, that some are concerned with this factor and as a consequence try to adapt their methods or to supply deficiencies. For example, the teacher who believes that children from ' poor ' homes (in the motivational or economic sense) need a period of browsing, of becoming familiar with books, of story periods intended to fire an interest and desire to learn to read, may delay formal instruction for the initial year of the infant course, whilst other teachers who consider children are there in the classroom to be taught, set out upon a teaching programme unrelated to considerations of background and expect children to learn to read from the very beginning of infant schooling.

There may be a definite relationship between teachers' attitudes to home background and their attitudes to what is termed ' reading readiness.' Until there are effective measures of teachers' attitudes to home background, there is no definite evidence as to the manner in which such attitudes can affect teachers' estimates of their children's abilities; e.g. the possibility of over- or under-estimation of abilities in relation to home background.

Conclusion

This very brief study does to some extent demonstrate that teachers' opinions reflect the results of measures of social reform operating in the community. Teachers in general probably have in mind differences in the emotional and motivational aspects of home background when they talk of a ' good ' or ' poor ' home. The bare-foot children of the thirties have been replaced by those who are emotionally deprived and/or whose parents are uninterested. Teachers' opinions reflect the changing emphasis within and the general impact of the socio-economic factor, and as Burt (1947) wrote thirteen years

ago " the very speed with which social characteristics, influences and conventions are changing means that psychologists who are engaged on research in the educational field must continually undertake intensive investigations on the sociological aspects of their problems." While we know so little about the extent and the effect of teachers' attitudes in regard to the socio-economic influences on children, the association between good attainment and socio-economic status should be interpreted with some caution.

Bibliography

BURT, C., " Selection for Secondary School." *British Journal of Educ. Psych.*, Vol. XVII, Part II, 1947.

FLEMING, C. M., " Socio-Economic Level and Test Performance." *British Journal of Educ. Psych.*, Vol. XIII, Part II, 1943.

FLOUD, J. E., HALSEY, A. H., and MARTIN, F. M., *Social Class and Educational Opportunity*. London: Heinemann, 1956.

FRASER, ELIZABETH, " Home Environment and the School." *Scottish Council for Research in Education*, 1959.

MORRIS, J. M., " Reading in the Primary School." *National Foundation for Educational Research* (Publication No. 12). London: Newnes, 1959.

Research Note:

THE LONG-TERM EFFECTS OF REMEDIAL TREATMENT

A Follow-up Enquiry based on a Case-study Approach[1]

by M. L. KELLMER PRINGLE

Department of Child Study, University of Birmingham.

THIS study is part of a larger investigation into the short-term and long-term effects of remedial education. It is based on the clinical work done by the Department of Child Study between the years 1948 and 1958. A letter was sent to 240 parents explaining that it was proposed to carry out a follow-up study of children who had received remedial treatment but that lack of staff prevented a personal interview; for this reason a written enquiry was being made about the child's progress since being discharged. The parents were told that we should like to obtain also a report from their child's present headmaster. The final response rate to the questionnaire[2] was 78%; in 14% of cases the family had moved and it proved impossible to obtain their new address; the remainder must be interpreted as a refusal to co-operate (8%). On the other hand, more co-operation than had been asked for was shown by over 10% of the parents; twenty of them sent letters in addition to completing the questionnaires and five telephoned to elaborate their answers to some of the questions.

To begin with, it was decided to analyse the questionnaires of those children who were still at school and who were discharged not earlier than the summer of 1952. There were 80 cases, 64 boys and 16 girls.

The main criteria governing the design of the questionnaire were brevity and simplicity: brief, factual questions minimise the difficulties of the less literate parents; moreover, most people dislike filling in questionnaires and thus the shorter and less complicated it looks, the more likely is it that it will be completed. Despite these precautions, our questionnaire (like most others) remains a coarse instrument and a second-best to follow-up interviews by a skilled Psychiatric Social Worker.

So far the questionnaires completed by the 80 parents and 66 head teachers have been analysed. The great majority of children were adolescents at the time of the follow-up enquiry and were attending secondary schools of various types.

Three of the questions put to the parents were of a very general nature: the first enquired whether the child was at the present time making better progress than he had done previously, or whether he was doing only as well as before or even worse; the second question aimed at assessing to what extent the parents felt (in retrospect) that the treatment given had helped the child; the third and most general question aimed at obtaining the parents' view regarding the child's present overall progress and adjustment, since they were asked whether he was causing them a great, some or no worry.

---

[1] A full account of this study will be published in *Vita Humana*, Vol. V, No. 1, 1962

[2] The questionnaire was drawn up and the method of analysing it was evolved by the writer in consultation with the Department's staff.

The great majority of parents consider that their child is making better educational progress since he received remedial treatment (Table I).

Table I

PRESENT SCHOLASTIC PROGRESS

N = 80

	Better than before	*As well as before*	*Worse than before*	*No reply*
N	63	10	4	3
% (*rounded figures*)	79	13	5	4

In about half the cases this treatment is thought to have helped a great deal while a small number felt it had not been of any help (Table II).

Table II

EXTENT TO WHICH TREATMENT HAS HELPED

N = 80

	A great deal	*To some extent*	*Not at all*
N	41	35	4
% (*rounded figures*)	51	44	5

Lastly the great majority of parents stated that their child gave them no cause for worry at the time of the enquiry (Table III).

Table III

PARENTS' PRESENT CONCERN REGARDING CHILD

N = 80

	No cause for worry	*Some cause for worry*	*A great deal of worry*
N	65	11	4
% *(rounded figures)*	81	14	5

Some more detailed questions were asked to elicit information on the educational progress each child was making at the time of the enquiry. Three descriptive categories were obtained. One third of the children are thought to be making excellent progress at school while about a quarter are still experiencing difficulties with one of the key subjects (Table IV).

Table IV

PRESENT EDUCATIONAL PROGRESS

N = 80

	Excellent	*Quite Satisfactory*	*Poor*
N	27	32	21
% *(rounded figures)*	34	40	26

Questions were also asked about present social and emotional development in view of the emphasis on this aspect of a child's growth during remedial treatment in our Department. In the great majority of cases, the parents consider their children's social and emotional development to be excellent or satisfactory. Very few fall into the lowest category of 'poor progress' (Table V).

Table V

PRESENT SOCIAL AND EMOTIONAL DEVELOPMENT
N = 80

	Excellent	*Satisfactory*	*Limited*	*Poor*
N	47	22	8	3
% (*rounded figures*)	59	28	10	4

Lastly, parental assessments of the child's present educational progress were tabulated in relation to their views of his present social and emotional development (Table VI).

Table VI

PRESENT EDUCATIONAL PROGRESS IN RELATION TO PRESENT SOCIAL AND EMOTIONAL DEVELOPMENT
N = 80

	Social and emotional development				
Educational progress	*Excellent*	*Satisfactory*	*Limited*	*Poor*	*Total*
Excellent ..	19	7	1	—	27
Satisfactory	19	10	3	—	32
Poor ..	9	5	4	3	21
Total ..	47	22	8	3	80

Development in both aspects is excellent in nearly a quarter of all cases, and poor in about 4%. If the excellent and satisfactory categories are taken together, 55 children fall within them for both educational progress and social and emotional development. Even of the children making poor educational progress, two-thirds show excellent or satisfactory social and emotional development.

There seems little doubt that the great majority of parents considered that remedial treatment had been of some help to their child. This view was held by most, even after more than six years had elapsed since regular attendance was terminated. The fact that most of the children did not give their parents any cause for concern is all the more satisfactory in view of the fact that most were adolescents at the time of the follow-up.

Our findings suggest that while children whose educational progress is poor may well be adjusted emotionally and socially, it is less likely that poor emotional and social development is accompanied by excellent educational progress.

Of course a study of this kind has many shortcomings. Since the children were discharged, a considerable period of time had elapsed, ranging from a minimum of eighteen months to six years. Obviously, all kinds of circumstances, both favourable and unfavourable, will have occurred in the children's lives; some of these may have helped their educational and social adjustment; others may well have undone any good that resulted from treatment in our department. We had no means of weighing up these circumstances. Another difficulty lies in the fact that we accepted the parents' judgements on their children's present progress and development, and relied on parental memory regarding the effect obtained by the treatment the children had received. On the other hand, the great majority of parents had been very concerned about their children's educational failure at the time when they were referred. One could expect therefore, that they would be able to compare the situation before and after remedial treatment had been given.

Lastly, there are as yet no reliable tools available to evaluate changes of attitude and personality, such as improvement in social relationships, in self-confidence, in perseverance and in motivation. For the time being we may well have to accept as valid criteria reports of improved behaviour or adjustment made by parents or teachers.

Bibliography:

AN ANNOTATED BIBLIOGRAPHY OF BRITISH RESEARCH ON TEACHING AND TEACHING ABILITY

by K. M. Evans
Lecturer and Research Officer, Faculty of Education, University College, Cardiff

SEVERAL bibliographies exist dealing with American research on teaching, but the results are not always directly applicable in this country, and it was thought that a corresponding bibliography of British work would be useful to researchers here as well as being of interest to those engaged in teaching and the training of teachers. It is not claimed that this list is exhaustive, but it is hoped that it is representative of the work which has been carried out. The accounts are arranged in two lists, one of articles in journals and the other of unpublished higher degree theses in university libraries. The order in both lists is chronological, as it was thought that this would help to show trends of research and bring out variations in the emphasis on topics over a period of years.

1. ARTICLES PUBLISHED IN JOURNALS.

Austin, F. M. (1931). **An analysis of the motives of adolescents for the choice of the teaching profession.**
British Journal of Educational Psychology, Vol. I, 87-103.

In an enquiry into the vocational choices of secondary school children teaching was the most popular profession, being chosen by 10 per cent. of the boys and 42 per cent. of the girls. Only a minority gave sound reasons for the choice or showed that they had any exact knowledge of careers or realised that they should choose one in which they had a good chance of success or happiness. The direct influence of relatives or teachers was the most frequent reason given for choosing teaching and there were many illustrations of a reluctance to grow up.

Cattell, R. B. (1931). **The assessment of teaching ability.**
British Journal of Educational Psychology, Vol. I, 48-72.

As a result of an enquiry among administrators, training college and university lecturers, teachers, students and pupils, a list of twenty-two qualities considered important in teachers was prepared. The relative importance of these qualities in young and mature teachers and men and women teachers was determined, and a rating scale for assessing student teachers was put forward. An abbreviated scale for use in the selection of candidates for training as teachers was also suggested.

PHILLIPS, M. (1931). **Professional courses in the training of teachers.**

British Journal of Educational Psychology, Vol. I, 225-245.

A questionnaire on the value of their courses of professional training was answered by 90 teachers who had been trained in various University Training Departments and two-year colleges. Many students, mainly women, considered that psychology was the most valuable part of the course, both personally and professionally. Teaching methods were regarded as of great importance by the men, most of whom were specialists. There was no overwhelming support for the view that the practical part of the course was more important than the theoretical or that training could be better conducted by schools than colleges. Teaching experience was not considered as a satisfactory alternative to training.

PHILLIPS, M. (1932). **Some problems of adjustment in the early years of a teacher's life.**

British Journal of Educational Psychology, Vol. II, 237-256.

Breakdowns among young teachers are usually the result of several factors rather than one single factor. They may be caused by personality difficulties, social or economic problems, or professional difficulties such as unsatisfactory placing of young teachers, bad working conditions, disciplinary difficulties, and unsatisfactory staff relationships. Of these, professional difficulties are possibly the most serious.

PINSENT, A. (1933). **Pre-college teaching experience and other factors in the teaching success of university students.**

British Journal of Educational Psychology, Vol. III, 109-126.

Among students in the University Training Department at Aberystwyth it was found that men, but not women, students with pre-college teaching experience tended to achieve higher Practical Teaching grades than those without such experience. Pre-college experience appeared to have an adverse effect upon the academic record of the women students but not on that of the men. Success in examinations in the Theory of Education did not seem to be connected with teaching experience either in the case of men or women.

VALENTINE, C. W. (1934). **An enquiry as to reasons for the choice of the teaching profession by university students.**

British Journal of Educational Psychology, Vol. IV, 237-259.

Among men students in four University Training Departments economic motives carried most weight in the choice of a teaching career. General liking for teaching and fondness for children were stronger motives among the women. Interest in favourite studies was a strong motive with both sexes. The influence of parents and teachers was not great. A minority of the students chose teaching as a stop-gap or because they saw no alternative. Objections to teaching included poor salaries, long training, small prospects of promotion, fear of getting into a rut, domination by examinations, and mental strain.

TURNBULL, G. H. (1934). **The influence of previous teaching experience on results obtained by students in a University Department of Education.**

British Journal of Educational Psychology, Vol. IV, 1-10.

This enquiry in the Sheffield University Training Department produced evidence that students with previous teaching experience did a little better in practical teaching but definitely worse in degree work than those who had no such experience. The performance of the two groups in educational theory was much the same. These results are compared with those obtained by Pinsent (1933).

DALDY, D. M. (1937). **A study of adaptability in a group of teachers.**

British Journal of Educational Psychology, Vol. VII, 1-22.

The success as teachers of a group of thirty-one domestic science students was studied in relation to their powers of adaptation. Adaptation was defined as involving the power to adjust to a situation without undue emotional conflict or display of emotion and without inducing conflict in the environment. Approximately half the group were classified as having difficulty in adaptation and this reacted unfavourably on their performance in teaching. Emotional relief obtained by discussing their difficulties resulted in an improvement in teaching.

LAWTON, J. A. (1939). **A study of factors useful in choosing candidates for the teaching profession.**

British Journal of Educational Psychology, Vol. IX, 131-144.

Academic ability, school records and interview impressions of training college students were correlated with their final teaching marks. All the coefficients obtained were significant. A composite estimate of teaching ability, based on academic attainment, interest in non-academic pursuits, and estimates of character qualities, was found to give a correlation with teaching marks sufficiently high to suggest that it would be a satisfactory instrument for the selection of students for admission to training colleges.

SAER, H. (1941). **A further investigation of pre-college teaching experience and other factors in the teaching success of university students.**

British Journal of Educational Psychology, Vol. XI, 183-196.

This study was carried out with a view to checking the results of the earlier enquiries by Pinsent (1933) and Turnbull (1934). On the whole the results obtained by the 1930-38 students at Aberystwyth corroborated Pinsent's findings. The effect of pre-college teaching experience on subsequent achievement appeared to be more clearly related to the type of experience and the stage at which it was gained than to its length.

WOOD, H. P. (1941). **The relative performance of Arts and Science graduates in a Teachers' Diploma examination.**

British Journal of Educational Psychology, Vol. XI, 8-19.

The performance in a postgraduate Teachers' Diploma examination of 304 Arts students was compared with that of 190 Science students. The performance of the Arts graduates was found to be significantly better than that of the Science graduates in all parts of the written examination, and better, but not significantly so, in practical teaching. Possible reasons for this are discussed.

TUDHOPE, W. B. (1942 and 1943). **A study of the training college final teaching mark as a criterion of future success in the teaching profession.**

British Journal of Educational Psychology, Vols. XII, 167-171 and XIII, 16-23.

In a group of 96 men and women teachers a high correlation (+0·81) was found between their college final teaching mark and an assessment made at least three years later by an H.M.I. Discrepancies appeared to be due to teachers teaching in schools or departments of different types from those for which they had been trained, to very abnormal conditions, to the attitude of the head teacher and the atmosphere of the school, to opportunities for specialising, and to emotional strain. These conditions in some cases raised and in others lowered the estimate.

TUDHOPE, W. B. (1944A). **Motives for the choice of the teaching profession by training college students.**

British Journal of Educational Psychology, Vol. XIV, 129-141.

Among 643 men and women training college students the most influential motives were a secure job, fondness for children, interest in a special subject, a desire to continue their own education, fondness for teaching, the possibility of doing good, and a good salary. Desirable motives outweighed undesirable ones. The decision to take up teaching was made earlier by women than by men, but only 7 per cent. of the whole group decided later than their first year in the sixth form.

TUDHOPE, W. B. (1944B). **Attitudes of secondary school authorities towards the training college course encountered by intending students.**

British Journal of Educational Psychology, Vol. XIV, 7-18.

In a survey of 643 men and women students in a mixed training college, 51 per cent. of the men and 71 per cent. of the women reported a generally encouraging attitude towards the college course from the heads of their secondary schools. Eleven per cent. of the men and 5 per cent. of the women reported a discouraging attitude and the remainder a non-committal one. Over 50 per cent. had been advised to go to a university instead. It is suggested that the attitude of secondary school authorities towards the training college course is unsatisfactory.

GRANT, P. J. T. (1950). **The social and educational background of emergency trained teachers and reasons for their choice of the profession.**

British Journal of Educational Psychology, Vol. XX, 164-173.

Information is given concerning the educational background, desire to teach, former employment, further education, decision to teach, pre-college training, hobbies and interests, and personalities of men and women in their late twenties and early thirties who were trained after the last war in four emergency training colleges.

EVANS, K. M. (1951). **A critical survey of methods of assessing teaching ability.**

British Journal of Educational Psychology, Vol XXI, 89-95.

An assessment of the value of pupil change, the judgement of experts, rating scales and pupil ratings as methods of evaluating teaching, together with evidence of the extent of agreement between these measures.

EVANS, K. M. (1952). **A study of attitude towards teaching as a career.**

British Journal of Educational Psychology, Vol. XXII, 63-69.

An outline of the results reported in the thesis by Evans (1946).

EVANS, K. M. (1953). **A further study of attitude towards teaching as a career.**

British Journal of Educational Psychology, Vol. XXIII, 58-63.

A test of attitude towards teaching as a career was administered to students in four training colleges and a University Training Department. The attitudes of the training college students were significantly more favourable than those of the university students, but in no case was there any significant correlation between attitude scores and Practical Teaching marks. In three of the colleges significant negative correlations were found between attitude scores and intelligence test results, and it is suggested that the reason for this result may be found in the social and economic conditions of the area from which most of their students were drawn.

SUTHERLAND, J. (1955). **A survey of students admitted to train as teachers in Scotland under the post-war emergency scheme.**

British Journal of Educational Psychology, Vol. XXV, 79-91.

All students in training for teaching in Scotland in the spring of 1949 sat an intelligence test and filled in a questionnaire. A number of these students had been admitted under the emergency scheme, and their results in their courses of study and the tests were later compared with those of the normal entrants. Little difference was found between the two groups.

OLIVER, J. N. (1956). **An analysis of the various factors associated with the assessment of teaching ability in physical education.** *British Journal of Educational Psychology*, Vol. XXVI, 66-77.

A rating scale for qualities important in teaching physical education was drawn up after consultation with men and women tutors and students, and was used in the assessment of the teaching of 127 men and 100 women students. Analysis of the results showed that three factors appeared to be important. These were a general teaching factor, a factor closely associated with personal and emotional qualities, and a factor associated with subject matter. There appeared to be general agreement throughout the profession as to the qualities and abilities necessary for successful teaching of physical education and there were no marked sex differences.

EVANS, K. M. (1957). **Is the concept of 'Interest' of significance to success in a teacher training course?** *Educational Review*, Vol. IX, 205-211.

A group of 41 women students in a University Training Department answered questionnaires about their general interests and their interest in various aspects of the work of a teacher. Comparison of their replies with their final examination marks in Practical Teaching and Theory of Education produced no evidence that students with many and varied interests were more successful in their course of training than those with fewer and less active interests. Neither was there any evidence that students who were keenly interested in the work of a teacher did better than those who were less interested. Significant negative correlations were obtained between interest in teaching and intelligence test results.

ROBERTSON, J. D. C. (1957). **An analysis of the views of supervisors on the attributes of successful graduate student teachers.** *British Journal of Educational Psychology*, Vol. XXVII, 115-126.

A brief account of the work reported in the thesis by Robertson (1953).

EVANS, K. M. (1958). **An examination of the Minnesota Teacher Attitude Inventory.** *British Journal of Educational Psychology*, Vol. XXVIII, 253-257.

At the beginning of their course, 109 students in a University Training Department were given the Moray House Adult Intelligence Test, Raven's Progressive Matrices (1937), the Minnesota Teacher Attitude Inventory, and a test of attitude towards teaching as a career. None of the test results gave significant correlations with Practical Teaching marks awarded at the end of the course, but all except the last gave correlations significant at the 5 per cent. level with marks in Theory of Education. The results of the M.T.A.I. did not correspond with either the American or Canadian norms, and a subsidiary experiment suggested that scores on the Inventory can be raised considerably by faking opinions where subjects wish to do this. Scores on the Inventory were significantly correlated with Progressive Matrices scores but not with Moray House Test results.

BURROUGHS, G. E. R. (1958). **A study of the interview in the selection of students for teacher training.**

British Journal of Educational Psychology, Vol. XXVIII, 37-46.

Students applying for admission to a university training department were interviewed and ratings made under a number of headings by the interviewers. Factor analysis of the ratings resulted in the identification of three factors concerned with the observable features of the applicant, his intellectual maturity and powers of self-expression, and personality traits thought to be important in teaching. The first two factors gave the best prediction of teaching ability. It is suggested that, within its own limitations, the interview is as good a predictor of success as most other measures.

COLLINS, M. (1959). **A follow-up study of some former graduate student teachers.**

British Journal of Educational Psychology, Vol. XXIX, 187-197.

The assessments made during training of 115 graduate student teachers were compared with assessments made one to seven years later by heads of schools in which they subsequently taught. It was found that 75 per cent. of the students achieved what would have been expected and a biserial correlation of +0·566 was obtained between heads' and tutors' assessments. Discrepancies between assessments seemed to show the influence of type of school, academic achievement, physical and mental health, family problems and length of professional experience on success in teaching.

CHARLTON, K., STEWART, W. A. C., and PAFFARD, M. K. (1958). **Students' attitudes to courses in Departments of Education in universities.**

British Journal of Educational Psychology, Vol. XXVIII, 243-252.

This is an interim report of a long-term study of the attitudes of students to their courses in University Departments of Education. It presents an analysis and comparison of the attitudes, as measured by a scale combining the Thurstone and Likert methods of construction, of students in the University of London and the University College of North Staffordshire.

CHARLTON, K., STEWART, W. A. C., and PAFFARD, M. K. (1960). **Students' attitudes to courses in Education.**

British Journal of Educational Studies, Vol. VIII, 148-164.

This is a later report of the 1958 study by the same authors (see above). On the whole the students expressed qualified approval of their Education courses and no evidence was found to support the sweeping negative criticisms sometimes voiced of the work of University Departments of Education.

II. UNPUBLISHED HIGHER DEGREE THESES.

PANTON, J. H. (1934). **The Assessment of Teaching Ability with Special Reference to Men Students in Training.**

Unpublished M.A. Thesis, University of London Library.

This investigation was aimed at discovering a method of standardising assessments of Practical Teaching and making them more reliable. Standards were found to vary from one college to another and distributions of teaching marks diverged from normal. Forty-nine men students were rated on a large number of qualities and the ratings were correlated with their teaching marks. The distribution of teaching marks was adjusted by using a rating scale based on the results.

HOLLIS, A. W. (1935). **The Personal Relationship in Teaching.**

Unpublished M.A. Thesis, University of Birmingham Library.

Over eight thousand school children between the ages of twelve and seventeen ranked seven qualities of teachers in order of preference. The results showed that they liked a friendly relationship with their teachers and that a close personal relationship with their teachers contributed to the solution of some of their difficulties. A liking for the teacher contributed to a liking for the subject taught. Unfair punishment was found to be useless and the pupils liked discipline based upon respect and liking for the teacher. A liking for wide interests in the teacher increased as the pupils got older.

EVANS, K. M. (1946). **A Study of Attitude towards Teaching as a Career.**

Unpublished M.A. Thesis, University of London Library.

A test of attitude towards teaching as a career was constructed and administered to 211 School Certificate candidates in eight grammar schools in England and Wales. Little difference was found between the attitudes of subjects in different parts of the country. Of a number of other variables considered, attitude to school seemed to be most closely correlated ($r = + 0{\cdot}36$) with attitude to teaching as a career. Achievement as measured by School Certificate results and attitude to working with the hands were both negatively correlated with attitude to teaching as a career. Neither the socio-economic level of the home nor the possession of relatives and/or friends who were teachers appeared to have any relationship to the attitude, but pupils whose interests were mainly academic or social had a significantly more favourable attitude than those whose interests were mainly practical.

THIMME-GOWDA, T. V. (1948) **A Study of the Attitudes of Teachers in England towards their Course of Training.**

Unpublished M.A. Thesis, University of London Library.

An attitude scale was drawn up dealing with different sections of the course of training.

1. Principles of Education: Theory and Practice.
2. Educational Psychology.

3. General Subjects (History, Geography, etc.).
4. Non-academic Subjects (P.E., Music, Art, etc.).
5. School Practice.

This scale was administered to 198 students in three two-year and one emergency training college and scores on it were intercorrelated with one another and with teaching success. Attitude to school practice correlated with teaching success to an extent of +0·208 (5% significance). Other attitude scores gave insignificant correlations with teaching success.

Attitudes to self, to society and to teaching work were also investigated. None of these correlated significantly with teaching success, though they did with some of the other attitude scores.

CHAMP, J. M. (1948). **A Study of the Attitude of Women Students, Teachers and Former Teachers towards Teaching as a Career.**

Unpublished M.A. Thesis, University of London Library.

An attitude scale given to 466 sixth form girls, training college, training department and emergency training college students, teachers and former teachers showed that the last was the only group having an unfavourable attitude to teaching and that there were no significant differences between the attitudes of teachers in different types of school. The most frequent causes given for withdrawal from the profession were marriage, overstrain due to weak discipline, other more attractive work, ill-health and retirement, and a dislike for dealing with people in masses. Salary and prospects of promotion were not mentioned. It is suggested that success sometimes depends on the conditions of work encountered and that apart from these the characteristics most likely to predispose to success in teaching are good intelligence, a good background, emotional stability and good health.

SKINNER, W. A. (1949). **An Investigation into Assessment of Teaching Ability in Teachers of Technical Subjects.**

Unpublished M.A. Thesis, University of London Library.

In any teaching situation there are three focal points to be considered, the teacher, the pupils, and the material arising during the activity. With this in mind a rating scale was drawn up and 87 students training for technical teaching were assessed on it. Correlations between scores on this rating scale and a global assessment of teaching ability were +0·88 when unweighted and +0·90 when weighted.

Factorial analysis was used and it was concluded that the ability to stimulate and arouse interest, the ability to evoke pupil co-operation, and the ability to make the lesson material of value to the pupils were those which made the highest contribution to teaching ability.

LOVELL, K. (1951). **An Investigation into Factors Underlying Teaching Ability in Primary and Secondary Schools with a View to Improving the Methods of Selecting Potential Teachers.**

Unpublished M.A. Thesis, University of London Library.

A large number of emergency training college students were assessed on twelve traits and the results correlated with Practical Teaching marks. Intercorrelations of the assessments showed a marked halo effect, and all correlations were positive. Factor analysis of the results showed that three factors were sufficient to account for teaching success. These were described as:

1. Intelligence and the willingness to use it in the education of children (57·1% of variance).
2. An empathy factor, the ability to appear live and interesting to children (5·6% of variance).
3. A speech factor (3·1% of variance).

The remaining 34·2% of the variance was contributed by specific factors.

In a group of students tested for the Cattell Primary Personality Factors of intelligence, cyclothymia, general emotional stability, dominance and surgency, it was found that students below the criterion level on all except dominance were likely to be average or below average teachers. There was no significant relationship between teaching marks and dominance.

BIRCHMORE, B. (1951). **A Study of the relationships between Pupils and Teachers in Certain Classes in a Secondary Grammar School.**

Unpublished M.A. Thesis, University of London Library.

This study aims at obtaining a greater understanding of the reasons why relationships go wrong between pupils and teachers. A list of qualities affecting the staff's pleasure in teaching individual boys, and a second list of qualities which the boys liked in their teachers were drawn up. The staff assessed all the boys according to the pleasure they obtained in teaching them and a group of 48 boys was chosen for more detailed study.

It was found that the boys' achievement rather than the effort they made contributed to the teachers' pleasure in teaching, and that there was a possible connection between the occupations of the boys' fathers and the staff's pleasure scores. The pupils held more complimentary opinions of the staff's qualities than the staff did of those of the pupils. The subject being taught appeared to affect pupil-teacher relations, and, where a teacher taught two subjects, relations appeared to be worse in the second subject.

Qualities which detracted most seriously from the staff's pleasure in teaching the boys were those such as laziness, disobedience, inattention and insolence which were liable to be interpreted as attacks upon the teacher or as leading to such attacks.

The thesis includes a number of case studies of individual pupils.

BURROUGHS, G. (1951). **Selection of Students for Training as Teachers.**
Unpublished Ph.D. Thesis, University of Birmingham Library.

An analysis was made of the results of using head teachers' reports, structured interviews and various intelligence and other tests in the selection of students for training as teachers.

It was found that estimates of teaching suitability were based mainly on non-cognitive qualities and that acceptability at school depended on social and athletic qualities. In the interview superficial qualities such as speech, manner and appearance were important and maturity and scholarship were minor factors. Using a schedule reduced the unreliability of the interview.

The interview results were affected by the adjustment of the candidate to the interviewers. The number on the panel was not important, but it was important that the candidate should be seen in as many situations as possible, so that different aspects of his behaviour could be observed.

For all practical purposes, tests, reports and interviews were unrelated and supplied different evidence about the candidates.

WADDINGTON, M. (1952). **An Investigation of the Effects of Contact with Children in Play Centre or Junior Club in Relation to the Training of Teachers.**
Unpublished M.A. Thesis, University of London Library.

In a survey of the use made of play centre or junior club work in the training of women teachers, questionnaires were filled in by 738 students of whom 479 helped in play centres and 259 did not. Individual interviews followed. Fifty-three per cent. of those who helped said that they found the work interesting and valuable because they had been able to observe children at play and therefore learn something of their interests.

In two of the colleges studied, students who had play centre experience did, in fact, improve their teaching mark more than those who had no such experience. A written examination paper showed no difference between the two groups. Centres allowing free play were more useful than those using formal, organised activities. The lecturers who used play centres in teacher training were enthusiastic about their value to the students.

UTTLEY, G. W. (1952). **A Study of Some Aspects of Personality in Relation to Teaching Ability for a Group of Students in an Emergency Training College.**
Unpublished M.A. Thesis, University of Birmingham Library.

The students in an emergency training college were asked to rank each other in order on leadership, co-operation, emotional stability, mental alertness, sympathy with tact, persistence, and expressiveness of personality. Positive correlations were found between all these traits and assessments of teaching ability made by tutors. In the case of leadership, mental alertness and emotional stability the correlations were significant.

A study of the best and worst teachers suggested that students who obtained high teaching marks also had high intelligence, but that high intelligence did not necessarily result in high teaching marks. For infant teachers other qualities were more important than intelligence.

There did not seem to be any connection between interests and teaching marks.

EVANS, K. M. (1952). **A Study of Teaching Ability at the Training College Stage in Relation to the Personality and Attitudes of the Student.**
Unpublished Ph.D. Thesis, University of London Library.

Pencil and paper tests designed to measure various aspects of resourcefulness, sociability, interests, interest in teaching and attitude to teaching as a career were administered to students in three training colleges and one University Training Department. Marked differences were found between the test results of students in different colleges and between the scores of sub-groups of students in one of the colleges, and it was clear that it was not possible to establish general norms but that separate norms would have to be established for each group of students.

When the test scores were correlated with Practical Teaching marks variations in the results were again found, and no test gave a significant level of correlation with all groups. Most of the correlation coefficients were small, but not smaller than those obtained in comparable aptitude studies using small, pre-selected groups of subjects. Attention is called to the contrast between these coefficients and the much higher ones frequently obtained when personality ratings made by college tutors, rather than objective test results, are correlated with teaching marks awarded by the same tutors.

PHILLIPS, A. S. (1953). **An Examination of Methods of Selection of Training College Students.**
Unpublished M.A. Thesis, University of London Library.

Tests of intelligence and English were applied to a group of training college students but the results showed little relationship to their teaching marks. A controlled projection test gave more promising results. Scores on a number of personality attributes were obtained from answers to this test, and three gave appreciable product-moment correlations with teaching marks. These qualities were called sympathetic understanding of children, friendliness and good disposition, and emotional stability. Their correlations with teaching marks were +0·27, +0·71, and +0·59 respectively. Their multiple correlation with teaching marks was +0·81.

ROBERTSON, J. D. C. (1953). **The Guidance and Assessment of the Practical Work of Graduate Student Teachers.**
Unpublished M.A. Thesis, University of London Library.

A list of 50 attributes considered to contribute to success in practice teaching was drawn up after consultation with eighteen supervisors. Analysis showed that these attributes could be arranged in eight groups, in order of importance, under the headings:

Attitude and insight in dealing with others.

Attitude and insight in learning to teach.

Teaching abilities.

Range of personality.

Attitudes as teacher.

Personality qualities: temperament and bearing.

Practical abilities.

Physical abilities.

It was found that the supervisors could be divided into three groups according to their points of view:

1. Supervisors who made special cognitive approaches to the attributes or who had special ways of perceiving teaching ability.
2. Supervisors who considered attributes associated with affective or interpersonal aspects particularly important, or who had special ways of responding to these aspects of teaching ability.
3. Supervisors who placed special emphasis on the process of developing teaching ability, or on conative aspects of learning to teach.

As a result of this work an experimental report sheet for rating graduate student teachers was drawn up, and its validity and reliability tested. These were considered to be satisfactory.

WALTERS, A. D. (1957). **An Investigation into the Value of Various Types of Information in the Selection of Training College Students and an Estimate of the Validity of Certain College Results.**

Unpublished M.A. Thesis, University of Liverpool Library.

The final examination marks and teaching assessments of students in three training colleges were compared with various data used in their selection. School certificate results gave the highest correlation with college examination marks, but their relationship to Practical Teaching marks was very variable. School estimates of intellectual ability were of doubtful value in predicting success but interview estimates of intellectual ability and personality were sometimes useful.

A follow-up study of some former students of one of the colleges showed that there was some correlation between professional success and success at college.

CORNWELL, J. (1958). **Sociometric Analysis in a Residential Training College.** Unpublished Ph.D. Thesis, University of London Library.

This thesis gives an account of the construction and validation of a near-sociometric test to predict the final performance, as measured by total course results, of men students in a two-year training college. As a result of experiments with three separate groups of students a test was evolved which could be administered early in the second term of a six term course. This test gave a rank correlation of $+0{\cdot}67 \pm{\cdot}064$ with final course results, and of $+0{\cdot}57 \pm{\cdot}078$ with final teaching marks. When the test was combined with examination results and first year teaching assessments the Pearson correlation with final results was $+0{\cdot}75 \pm{\cdot}048$.

It is claimed that this test identifies early in the course those students whose final performance is likely to be weak. It also is said to differentiate between students who are likely to fail the course and who should be advised to leave, and those, who, though weak may be expected to pass if given help at the onset of their difficulties.

©

National Foundation for Educational Research in England and Wales
Printed by King, Thorne & Stace Limited, 1-13 Jubilee Street, Brighton, 1
and published by Newnes Educational Publishing Co., Ltd., Tower House, Southampton Street, London, W.C.2.,
for The National Foundation for Educational Research in England and Wales, 79 Wimpole Street, London, W.1.

Publications of the

NATIONAL FOUNDATION FOR EDUCATIONAL RESEARCH

BOOKS

1. A List of Researches in Education and Educational Psychology. 1918-48. By A. M. Blackwell, M.A., B.Sc. (out of print).

2. The Allocation of Primary School Leavers to Courses of Secondary Education. First Interim Report. By A. F. Watts, M.A., D.Lit., and Patrick Slater, M.A., 7s. 6d.

3. A Survey of Rewards and Punishments in Schools. Based on Researches carried out by M. E. Highfield, M.A., Ph.D., and A. Pinsent, M.B.E., M.A., B.Sc., £2 2s. 0d.

4. The Educability of Cerebral Palsied Children. By M. I. Dunsdon, M.A., F.B.Ps.S., £1 1s. 0d.

5. A Second List of Researches in Education and Educational Psychology, 1949-51. By A. M. Blackwell, M.A., B.Sc., £1 1s. 0d.

6. Secondary School Entrance Examinations. By A. F. Watts and others, 6s. 6d. net, (cloth cased), 3s. 6d. (limp covers).

7. Lists of Researches in Education and Educational Psychology. Supplement I. 1952-53. By A. M. Blackwell, M.A., B.Sc. (out of print).

8. Pupils' School Records. By Alice S. Walker, B.Sc., £1 1s. 0d.

9. Lists of Researches in Education and Educational Psychology. Supplement II. 1954-55. By A. M. Blackwell, M.A., B.Sc., 15s.

10. Admission to Grammar Schools. Third Interim Report on the Allocation of Primary School Leavers to Courses of Secondary Education. By Alfred Yates, B.A., B.Ed., and D. A. Pidgeon, B.Sc., £1 5s. 0d.

11. Lists of Researches in Education and Educational Psychology. Supplement III. 1956-57. By A. M. Blackwell, M.A., B.Sc., 15s.

12. Reading in the Primary School. By Joyce M. Morris, B.A., Ph.D., with a Statistical Appendix by P. M. Grundy, B.A., Ph.D., 15s. net.

Orders for the Foundation's Publications should be placed with its Publishers—

Newnes Educational Publishing Co., Ltd.
Tower House, Southampton Street, London, W.C.2

CONTENTS

PAGE

Articles

THE MEASUREMENT OF PERSONALITY—I. *F. W. Warburton* 2

CRIME, CAUSE AND TREATMENT: RECENT RESEARCH AND THEORY. *Leslie T. Wilkins* 18

EDUCATIONAL GUIDANCE: THE DESIGN, CONSTRUCTION AND USE OF STANDARDISED TESTS—PART 2: THE INTERPRETATION OF TEST SCORES. *D. A. Pidgeon* .. 33

Research Notes

TEACHERS' VIEWS ON PRIMARY SCHOOL STREAMING. *Brian Jackson* 44

OVER 135 I.Q. *John Nisbet and Alastair Gammie* 53

TEACHERS AND THE SOCIO-ECONOMIC FACTOR. *E. J. Goodacre* 56

THE LONG-TERM EFFECTS OF REMEDIAL TREATMENT: A FOLLOW-UP ENQUIRY BASED ON A CASE-STUDY APPROACH. *M. L. Kellmer Pringle* 62

Bibliography

AN ANNOTATED BIBLIOGRAPHY OF BRITISH RESEARCH ON TEACHING AND TEACHING ABILITY. *K. M. Evans* 67